BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Europäisches Patentamt	European Patent Office	Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

97200831.2

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 97200831.2
Demande no:

Anmeldetag:
Date of filing: 19.03.97
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

MOGEN INTERNATIONAL N.V.
Einsteinweg 97
NL-2333 CB Leiden
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Antifungal proteins, dna coding therefor, and hosts incorporating same

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C07K14/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

NL

ANTIFUNGAL PROTEINS, DNA CODING THEREFORE, AND HOSTS INCORPORATING SAME

FIELD OF THE INVENTION

The present invention relates to antifungal proteins, DNA coding therefor and hosts incorporating the DNA, as well as methods of combating fungal pathogens by causing said fungal pathogens to be contacted with said protein or proteins.

The invention further relates to plants, incorporating and expressing DNA coding for antifungal proteins, and to plants which as a result thereof show reduced susceptibility to fungal pathogens, in particular to the Oomycetes *Phytophthora* and *Pythium*.

BACKGROUND ART

Phytophthora infestans belongs to the group of fungi referred to as Oomycetes. *Phytophthora infestans* infects various members of Solanaceae, such as potato, tomato and some ornamentals. It causes late blight of potatoes and tomatoes affecting all parts except roots. Geographically, the fungus is widely distributed, and it can be found in all potato-producing countries. Economically late blight in potatoes is of major importance, as infection early in the season can severely reduce crop yield. Currently the disease is controlled by spraying chemical fungicides (dithiocarbamates, such as mancozeb, manec and zineb) regularly. Both from an environmental and economical point of view, biological control of diseases caused by *Phytophthora infestans* could have advantages over the use of chemical fungicides.

Pythium also belongs to the group of fungi referred to as Oomycetes. The genus *Pythium* differs from the related genus *Phytophthora* by forming relatively undifferentiated sporangia. Geographically, this fungus is widely distributed on all continents. The first main type of disease caused by *Pythium* species is damping-off, due to sudden and fast developing attacks on young seedlings in the field or in nurseries. *Pythium* species cause a second type of disease which is root necrosis and causes a general slowing of plant growth (for example wheat and maize) and loss of yield.

The main losses caused by *Pythium* in Europe are to field crops such as sugarbeet. In principle, losses tend to be all-or-nothing. Similarly, nursery sowings of ornamentals and forest trees may be completely

destroyed. (For a review on Oomycetes, *vide*: European Handbook of Plant Diseases, ed. by I.M. Smith *et al.*, 1988, Blackwell Scientific Publications, Ch.8)

A protein with antifungal activity, isolated from TMV-induced tobacco leaves, which is capable of causing lysis of germinating spores and hyphal tips of *Phytophthora infestans* and which causes the hyphae to grow at a reduced rate, was disclosed in WO91/18984 A1. This protein has an apparent molecular weight of about 24 kDa and was named AP24. Comparison of its complete amino acid sequence, as deduced from the nucleic acid sequence of the AP24 gene, with proteins known from databases revealed that the protein was an osmotin-like protein.

Despite initial success in combating fungal pathogens, such as *Phytophthora infestans*, and the genetic engineering of plants capable of producing these antifungal proteins with activity against this fungal pathogen there remains a need to identify and isolate other proteins with antifungal activity against this fungus.

SUMMARY OF THE INVENTION

The present invention provides an isolated protein obtainable from a plant source which has anti- fungal activity, especially directed to Oomycetes, and preferably to *Phytophthora* and/or *Pythium* and a molecular weight of about 59 ± 5 kDa as judged by SDS PAGE-electrophoresis. A more preferred protein is one that is obtainable from sunflower plants. An even more preferred protein is one that is obtainable from sunflower leaves induced with sodium salicylate. A still more preferred isolated protein is characterised in that it is selected from the group of proteins having the amino acid sequence selected from the group comprising of the amino acid sequences depicted in SEQ ID NO's: 1, 2 or 6, 16 or 20, as well as muteins thereof which have antifungal and especially anti-*Phytophthora* and/or anti-*Pythium* activity. A still further preferred protein according to the invention is one characterised in that it comprises a protein thatcomprises the amino acid sequence as represented by SEQ ID NO: 16 or SEQ ID NO: 20, or by a part of said sequence like represented in SEQ ID NO: 6.

The invention also embraces an isolated DNA sequence comprising an open reading frame capable of encoding a protein according to the invention, preferably characterised in that the open reading frame is

capable of encoding a protein according to the invention, and DNA capable of hybridising therewith under stringent conditions.

The invention also provides a chimeric DNA sequence according to the invention further comprising a transcriptional initiation region and, optionally, a transcriptional termination region, so linked to said open reading frame as to enable the DNA to be transcribed in a living host cell when present therein, thereby producing RNA which comprises said open reading frame. A preferred chimeric DNA sequence according to the invention is one, wherein the RNA comprising said open reading frame is capable of being translated into protein in said host cell, when present therein, thereby producing said protein. Especially preferred are DNA sequences comprising a sequence as depicted in SEQ ID NO: 15 or SEQ ID NO: 19.

The invention also embraces a chimeric DNA sequence comprising a DNA sequence according to the invention, which may be selected from replicons, such as bacterial cloning plasmids and vectors, such as a bacterial expression vector, a (non-integrative) plant viral vector, a Ti-plasmid vector of *Agrobacterium*, such as a binary vector, and the like, as well as a host cell comprising a replicon or vector according to the invention, and which is capable of maintaining said replicon once present therein. Preferred according to that embodiment is a host cell which is a plant cell, said vector being a non-integrative viral vector.

The invention further provides a host cell stably incorporating in its genome a chimeric DNA sequence according to the invention, such as a plant cell, as well as multicellular hosts comprising such cells, or essentially consisting of such cells, such as plants. Especially preferred are plants characterised in that the chimeric DNA according to the invention is expressed in at least a number of the plant's cells causing the said antifungal protein to be produced therein.

According to yet another embodiment of the invention a method for producing a protein with antifungal activity is provided, characterised in that a host cell according to the invention is grown under conditions allowing the said protein to be produced by said host cell, optionally followed by the step of recovering the protein from the host cells.

Another part of the invention is directed to the antifungal use of a protein which has reticuline oxidase activity.

The invention provides also for the use of a protein according to the invention for retarding the growth of fungi, preferably Oomycetes and more preferably *Phytophthora* and *Pythium*, preferably characterised in that spores of the said fungus are caused to be contacted with said protein. According to yet another embodiment, retarding the growth of the fungus *Phytophthora* and/or *Pythium* is on plant leaves, characterised in that hyphae thereof, or spores thereof, are caused to be contacted with a protein produced from a host cell according to the invention capable thereof.

The invention also provides a method for obtaining plants with reduced susceptibility to fungi, especially *Phytophthora* and/or *Pythium*, comprising the steps of

(a) introducing into ancestor cells which are susceptible of regeneration into a whole plant,

- a chimeric DNA sequence comprising an open reading frame capable of encoding a protein according to claim 1, said open reading frame being operatively linked to a transcriptional and translational region and, optionally, a transcriptional termination region, allowing the said protein to be produced in a plant cell that is susceptible to infection by said fungus, and
- a chimeric DNA sequence capable of encoding a plant selectable marker allowing selection of transformed ancestor cells when said selectable marker is present therein, and

(b) regenerating said ancestor cells into plants under conditions favouring ancestor cells which have the said selectable marker, and

(c) identifying a plant which produces a protein according to claim 1, thereby reducing the susceptibility of said plant to infection by said fungus.

Preferred according to the invention is a method characterised in that step (a) is performed using an *Agrobacterium tumefaciens* strain capable of T-DNA transfer to plant cells and which harbours the said chimeric DNA cloned into binary vector pMOG800; another preferred method is when step (b) is performed in the presence of an antibiotic favouring cells which have a neomycin phosphotransferase.

The invention further provides an antifungal composition comprising a protein according to the invention and a suitable carrier.

An antibody, capable of reacting with an N-terminal fragment of

a protein according to the invention, preferably to the peptide represented by SEQ ID NO: 6, 16 or 20, is also provided. The antibody is suitably used to detect expression levels of chimeric DNA according to the invention in host cells and multicellular hosts, preferably plants, capable of producing a protein according to the invention. The invention also provides a nucleic acid sequence obtainable from a gene encoding a protein according to the invention, said nucleic acid sequence having tissue-specific transcriptional regulatory activity in a plant. The invention specifically provides a nucleic acid sequence obtainable from the region upstream of the translational initiation site of said gene, preferably at least 500 nucleotides immediately upstream of the translational initiation site of said gene.

DESCRIPTION OF THE FIGURES

Figure 1: SDS-PAGE (12.5%) of the different purification steps of MS59 sunflower protein. Mw= molecular weight markers; 1= crude sunflower protein extract after gel filtration (G25); 2= protein fraction bound to cation exchange chromatography (S-sepharose); 3= pool of active fractions after cation exchange chromatography (Mono S); 4= flow through from hydrophobic interaction chromatography (phenyl superose); 5= active fractions after gel filtration.

Figure 2: SDS-PAGE (12.5%) of different fractions (number 6 to 16) of the gelfiltration (SD75) column. Fraction 10 to 15 was tested in 3 dilutions for growth inhibition on *Phytophthora infestans* (PANEL A) and on *Pythium ultimum* (PANEL B)

Figure 3: SDS-PAGE (12.5%) of fractions eluted from nine gel slices (lane 1 to 9) of a native PAGE in which a MS59 containing SD75 fraction (SD75 fraction 13)was separated. Right panel: SDS-PAGE (12.5%) with SD75 fraction 13 (L) and two fractions of elution experiment fraction 2 (with MS59) and fraction 5 (with a ~30 kD protein). Bottom panel: growth inhibition of *Phytophthora infestans* tested with elution fraction 1 to 6, with 5 µl and 1 µl added per well.

Figure 4: Microscopical analysis of an *in vitro* fungal inhibition assay 24 hours after addition of *Phytophthora infestans* zoosporangia to PDA medium. Left panel: control incubation, only MES buffer was added. Right panel:*E. coli*-produced MS59 in MES buffer was added to the incubation.

Figure 5: Microscopical analysis of an *in vitro* fungal inhibition assay 24 hours after addition of *Pythium ultimum* hyphal fragments to PDA medium. Left panel: control incubation, only MES buffer was added. Right panel: *E. coli*-produced MS59 in MES buffer was added to the incubation.

DETAILED DESCRIPTION OF THE INVENTION

The antifungal effect of the protein(s) of the invention has been demonstrated in *in vitro* assays for the following fungi; *Phytophthora infestans, Phytophthora cactorum, Phytophtora nicotiana, Phytophthora megasperma, Pythium ultimum, Pythium sylvaticum, Pythium violae* and for *Pythium paroecandrum* for purposes of illustration. It will be clear, that the use of the protein(s) of the invention, or DNA encoding therefore, for use in a process of combating fungi is not limited to the mentioned fungi. There is no reason to assume that the protein(s) according to the invention do not possess antifungal activity against a far broader range of fungi than those tested here, especially in the class of Oomycetes.

Although the invention is illustrated in detail for transgenic tomato, tobacco, carrot, potato and *Brassica napus* plants, it should be understood that any plant species that is subject to some form of fungal attack, especially from the fungi mentioned above, may be provided with one or more plant expressible gene constructs, which when expressed overproduce the protein(s) of the invention in said plant in order to decrease the rate of infectivity and/or the effects of such attack. The invention can even be practiced in plant species that are presently not amenable for transformation, as the amenability of such species is just a matter of time and because transformation as such is of no relevance for the principles underlying the invention. Hence, plants for the purpose of this description shall include angiosperms as well as gymnosperms, monocotyledonous as well as dicotyledonous plants, be they for feed, food or industrial processing

purposes; included are plants used for any agricultural or horticultural purpose including forestry and flower culture, as well as home gardening or indoor gardening, or other decorative purposes.

The protein according to the present invention may be obtained by isolating it from any suitable plant source material containing it. A particularly suitable source comprises leaves of the sunflower (*Helianthus*). The presence of antifungal proteins according to the invention in plant source material can readily be determined for any plant species by making plant extracts from those species and testing those extracts for the presence of antifungal activity using an *in vitro* antifungal assay as described herein, further fractionating the obtained samples by any suitable protein fractionation technique in conjunction with the *in vitro* assay until an antifungal fraction is obtained which comprises an approximately 59 kDa protein, internally denoted as MS59, which in isolated form shows antifungal activity. Especially, fractions may be tested for antifungal activity on Oomycetes, for example, *Phythophthora* or *Pythium ultimum* and the like, or other fungi, such as the Basidiomycetes, Ascomycetes, Zygomycetes or other classes or subclasses.

Alternatively, antifungal proteins according to the invention may be obtained by cloning DNA comprising an open reading frame capable of encoding said protein, or the precursor thereof, linking said open reading frame to a transcriptional, and optionally a translational initiation and transcriptional termination region, inserting said DNA into a suitable host cell and allowing said host cell to produce said protein. Subsequently, the protein may be recovered from said host cells, preferably after secretion of the protein into the culture medium by said host cells. Alternatively, said host cells may be used directly in a process of combating fungal pathogens according to the invention as a pesticidal acceptable composition.

Host cells suitable for use in a process of obtaining a protein according to the invention may be selected from prokaryotic microbial hosts, such as bacteria *e.g. Agrobacterium*, *Bacillus*, Cyanobacteria, *E.coli* , *Pseudomonas*, and the like, as well as eukaryotic hosts including yeasts, *e.g. Saccharomyces cerevisiae*, fungi, *e.g. Trichoderma* and plant cells, including protoplasts.

In a method of retarding the growth of the fungus *Phytophthora infestans* on plant leaves, characterised in that hyphae thereof, or spores thereof, are caused to be contacted with a protein produced from a host cell, host cells may suitably be selected from any species routinely used as biological fungicides.

Although the invention is set out in more detail using *Phytophthora infestans* and *Pythium* as an example, it will be clear that proteins according to the invention may be tested for antifungal activity other than anti-*Phytophthora* activity using an antifungal assay similar to that described in the present specification. Suitable antifungal assays have been described for several other fungi in European patent application 440 304 A1.

The word protein means a sequence of amino acids connected trough peptide bonds. Polypeptides or peptides are also considered to be proteins. Muteins of the protein of the invention are proteins that are obtained from the proteins depicted in the sequence listing by replacing, adding and/or deleting one or more amino acids, while still retaining their antifungal activity. Such muteins can readily be made by protein engineering *in vivo*, e.g. by changing the open reading frame capable of encoding the antifungal protein such that the amino acid sequence is thereby affected. As long as the changes in the amino acid sequences do not altogether abolish the antifungal activity such muteins are embraced in the present invention.

The present invention provides a chimeric DNA sequence which comprises an open reading frame capable of encoding a protein according to the invention. The expression chimeric DNA sequence shall mean to comprise any DNA sequence which comprises DNA sequences not naturally found in nature. For instance, chimeric DNA shall mean to comprise DNA comprising the said open reading frame in a non-natural location of the plant genome, notwithstanding the fact that said plant genome normally contains a copy of the said open reading frame in its natural chromosomal location. Similarly, the said open reading frame may be incorporated in the plant genome wherein it is not naturally found, or in a replicon or vector where it is not naturally found, such as a bacterial plasmid or a viral vector. Chimeric DNA shall not be limited to DNA molecules which are replicable in a host, but shall also mean to comprise DNA capable of being ligated into a replicon, for instance by virtue of specific adaptor sequences, physically

linked to the open reading frame according to the invention. The open reading frame may or may not be linked to its natural upstream and downstream regulatory elements.

The open reading frame may be derived from a genomic library. In this latter it may contain one or more introns separating the exons making up the open reading frame that encodes a protein according to the invention. The open reading frame may also be encoded by one uninterrupted exon, or by a cDNA to the mRNA encoding a protein according to the invention. Open reading frames according to the invention also comprise those in which one or more introns have been artificially removed or added. Each of these variants is embraced by the present invention.

Also part of the invention are chimeric DNA sequences coding for an antifungal protein which comprise one or more of the EST-sequences shown in SEQ ID NO's: 21 to 47. As can be derived from the sequence listings these EST's for which no function was hitherto known share a considerable homology with the DNA sequence coding for the protein isolated from *Helianthus*.

Another source of homologous sequences is found in the class of reticuline oxidases (EC 1.5.3.9) which are known from the biochemical pathway for the production of sanguinarine in *Papaver somniferum* and scoulerine in *Eschscholtzia californica* (Kutchan, T.M. and Dittrich M., J. Biol. Chem. 270(41), 24475-24481, 1995; Facchini, P.J. *et al.*, Plant Physiol. 112, 1669-1677, 1996). Although it is known that fungal elicitors are capable of the induction of sanguinarine (Facchini, P.J. *et al.*, Plant Physiol. 111, 687-697, 1996) and that also the reticuline oxidase itself is induced by an elicitor (Dittrich, M. and Kutchan, T.M., Proc. Natl. Acad. Sci. USA 88, 9969, 9973, 1991) it is surprising that the enzyme itself has such strong antifungal properties, especially since sanguinarine is known to have antimicrobial activities.

In order to be capable of being expressed in a host cell a chimeric DNA according to the invention will usually be provided with regulatory elements enabling it to be recognised by the biochemical machinery of the host and allowing for the open reading frame to be transcribed and/or translated in the host. It will usually comprise a transcriptional initiation region which may be suitably derived from

any gene capable of being expressed in the host cell of choice, as well as a translational initiation region for ribosome recognition and attachment. In eukaryotic cells, an expression cassette usually comprises in addition a transcriptional termination region located downstream of said open reading frame, allowing transcription to terminate and polyadenylation of the primary transcript to occur. In addition, the codon usage may be adapted to accepted codon usage of the host of choice. The principles governing the expression of a chimeric DNA construct in a chosen host cell are commonly understood by those of ordinary skill in the art and the construction of expressible chimeric DNA constructs is now routine for any sort of host cell, be it prokaryotic or eukaryotic.

In order for the open reading frame to be maintained in a host cell it will usually provided be in the form of a replicon comprising said open reading frame according to the invention linked to DNA which is recognised and replicated by the chosen host cell. Accordingly the selection of the replicon is determined largely by the host cell of choice. Such principles as govern the selection of suitable replicons for a particular chosen host are well within the realm of the ordinary skilled person in the art.

A special type of replicon is one capable of transferring itself, or a part thereof, to another host cell, such as a plant cell, thereby co-transferring the open reading frame according to the invention to said plant cell. Replicons with such capability are herein referred to as vectors. An example of such vector is a Ti-plasmid vector which, when present in a suitable host, such as *Agrobacterium tumefaciens*, is capable of transferring part of itself, the so-called T-region, to a plant cell. Different types of Ti-plasmid vectors (*vide:* EP 0 116 718 B1) are now routinely being used to transfer chimeric DNA sequences into plant cells, or protoplasts, from which new plants may be generated which stably incorporate said chimeric DNA in their genomes. A particularly preferred form of Ti-plasmid vectors are the so-called binary vectors as claimed in (EP 0 120 516 B1 and US 4,940,838). Other suitable vectors, which may be used to introduce DNA according to the invention into a plant host, may be selected from the viral vectors, *e.g.* non-integrative plant viral vectors, such as derivable from the double stranded plant viruses (*e.g.* CaMV) and single stranded viruses, gemini viruses and

the like. The use of such vectors may be advantageous, particularly when it is difficult to stably transform the plant host. Such may be the case with woody species, especially trees and vines.

The expression "host cells incorporating a chimeric DNA sequence according to the invention in their genome" shall mean to comprise cells, as well as multicellular organisms comprising such cells, or essentially consisting of such cells, which stably incorporate said chimeric DNA into their genome thereby maintaining the chimeric DNA, and preferably transmitting a copy of such chimeric DNA to progeny cells, be it through mitosis or meiosis. According to a preferred embodiment of the invention plants are provided, which essentially consist of cells which incorporate one or more copies of said chimeric DNA into their genome, and which are capable of transmitting a copy or copies to their progeny, preferably in a Mendelian fashion. By virtue of the transcription and translation of the chimeric DNA according to the invention in some or all of the plant's cells, those cells as produce the antifungal protein will show enhanced resistance to fungal infections, especially to *Phytophthora* infections. Although the principles as govern transcription of DNA in plant cells are not always understood, the creation of chimeric DNA capable of being expressed in substantially a constitutive fashion, that is, in substantially most cell types of the plant and substantially without serious temporal and/or developmental restrictions, is now routine. Transcription initiation regions routinely in use for that purpose are promoters obtainable from the cauliflower mosaic virus, notably the 35S RNA and 19S RNA transcript promoters and the so-called T-DNA promoters of *Agrobacterium tumefaciens*, in particular to be mentioned are the nopaline synthase promoter, octopine synthase promoter (as disclosed in EP 0 122 791 B1) and the mannopine synthase promoter. In addition plant promoters may be used, which may be substantially constitutive, such as the rice actin gene promoter, or *e.g.* organ-specific, such as the root-specific promoter. Alternatively, pathogen-inducible promoters may be used such as the PRP1 promoter (also named gst1 promoter) obtainable from potato (Martini N. *et al.* (1993), Mol. Gen. Genet. 263, 179-186). The choice of the promoter is not essential, although it must be said that constitutive high-level promoters are slightly preferred. It is further known that duplication of certain elements, so-called enhancers, may considerably enhance the

expression level of the DNA under its regime (*vide* for instance: Kay R. *et al.* (1987), Science 236, 1299-1302: the duplication of the sequence between -343 and -90 of the CaMV 35S promoter increases the activity of that promoter). In addition to the 35S promoter, singly or doubly enhanced, examples of high-level promoters are the light-inducible ribulose bisphosphate carboxylase small subunit (*rbc*SSU) promoter and the chlorophyll a/b binding protein (Cab) promoter. Also envisaged by the present invention are hybrid promoters, which comprise elements of different promoter regions physically linked. A well known example thereof is the so-called CaMV enhanced mannopine synthase promoter (US Patent 5,106,739), which comprises elements of the mannopine synthase promoter linked to the CaMV enhancer.

As regards the necessity of a transcriptional terminator region, it is generally believed that such a region enhances the reliability as well as the efficiency of transcription in plant cells. Use thereof is therefore strongly preferred in the context of the present invention.

Another aspect of gene expression in transgenic plants concerns the targeting of antifungal proteins to the extracellular space (apoplast). Naturally intracellularly occurring proteins, among which proteins according to the present invention, may be caused to be targeted to the apoplast by removal of the C-terminal propeptide, *e.g.* by modifying the open reading frame at its 3' end such that protein is caused to be C-terminally truncated. A certain number of amino acids of the C-terminal part of the protein was found to be responsible for targeting of the protein to the vacuole (*e.g. vide* WO91/18984 A1). By introducing a translational stopcodon in the open reading frame the truncated protein is caused to be targeted to the apoplast.

As regards the applicability of the invention in different plant species, it has to be mentioned that one particular embodiment of the invention is merely illustrated with transgenic tomato and tobacco plants as an example, the actual applicability being in fact not limited to these plant species. Any plant species that is subject to some form of fungal attack, in particular by Oomycetes such as *Phytophthora infestans*, may be treated with proteins according to the invention, or preferably, be provided with a chimeric DNA sequence according to the invention, allowing the protein to be produced in

some or all of the plant's cells.

Although some of the embodiments of the invention may not be practicable at present, *e.g.* because some plant species are as yet recalcitrant to genetic transformation, the practicing of the invention in such plant species is merely a matter of time and not a matter of principle, because the amenability to genetic transformation as such is of no relevance to the underlying embodiment of the invention.

Transformation of plant species is now routine for an impressive number of plant species, including both the *Dicotyledoneae* as well as the *Monocotyledoneae*. In principle any transformation method may be used to introduce chimeric DNA according to the invention into a suitable ancestor cell, as long as the cells are capable of being regenerated into whole plants. Methods may suitably be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. *et al.*, 1982, Nature 296, 72-74; Negrutiu I. *et al*, June 1987, Plant Mol. Biol. 8, 363-373), electroporation of protoplasts (Shillito R.D. *et al.*, 1985 Bio/Technol. 3, 1099-1102), microinjection into plant material (Crossway A. *et al.*, 1986, Mol. Gen. Genet. 202, 179-185), (DNA or RNA-coated) particle bombardment of various plant material (Klein T.M. *et al.*, 1987, Nature 327, 70), infection with (non-integrative) viruses and the like. A preferred method according to the invention comprises *Agrobacterium*-mediated DNA transfer. Especially preferred is the use of the so-called binary vector technology as disclosed in EP A 120 516 and U.S. Patent 4,940,838).
Tomato transformation is preferably done essentially as described by Van Roekel et al. (Van Roekel, J.S.C., Damm, B., Melchers, L.S., Hoekema, A. (1993). Factors influencing transformation frequency of tomato (*Lycopersicon esculentum*). Plant Cell Reports, 12, 644-647).
Potato transformation is preferably done essentially as described by Hoekema et al. (Hoekema, A., Huisman, M.J., Molendijk, L., van den Elzen, P.J.M., and Cornelissen, B.J.C. (1989). The genetic engineering of two commercial potato cultivars for resistance to potato virus X. Bio/Technology 7, 273-278).
Generally, after transformation plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant expressible genes co-transferred with the nucleic acid sequence encoding the protein according to the invention, whereafter the

transformed material is regenerated into a whole plant.

Although considered somewhat more recalcitrant towards genetic transformation, monocotyledonous plants are amenable to transformation and fertile transgenic plants can be regenerated from transformed cells or embryos, or other plant material. Presently, preferred methods for transformation of monocots are microprojectile bombardment of embryos, explants or suspension cells, and direct DNA uptake or electroporation (Shimamoto, *et al*, 1989, Nature 338, 274-276). Transgenic maize plants have been obtained by introducing the *Streptomyces hygroscopicus bar*-gene, which encodes phosphinothricin acetyltransferase (an enzyme which inactivates the herbicide phosphinothricin), into embryogenic cells of a maize suspension culture by microprojectile bombardment (Gordon-Kamm, 1990, Plant Cell, 2, 603-618). The introduction of genetic material into aleurone protoplasts of other monocot crops such as wheat and barley has been reported (Lee, 1989, Plant Mol. Biol. 13, 21-30). Wheat plants have been regenerated from embryogenic suspension culture by selecting only the aged compact and nodular embryogenic callus tissues for the establishment of the embryogenic suspension cultures (Vasil, 1990 Bio/Technol. 8, 429-434). The combination with transformation systems for these crops enables the application of the present invention to monocots.

Monocotyledonous plants, including commercially important crops such as rice and corn are also amenable to DNA transfer by *Agrobacterium* strains (*vide* WO 94/00977; EP 0 159 418 B1; Gould J, Michael D, Hasegawa O, Ulian EC, Peterson G, Smith RH, (1991) Plant. Physiol. 95, 426-434).

Following DNA transfer and regeneration, putatively transformed plants may be evaluated, for instance using Southern analysis, for the presence of the chimeric DNA according to the invention, copy number and/or genomic organization. In addition, or alternatively, expression levels of the newly introduced DNA may be undertaken, using Northern and/or Western analysis, techniques well known to persons having ordinary skill in the art. After the initial analysis, which is optional, transformed plants showing the desired copy number and expression level of the newly introduced chimeric DNA according to the invention may be tested for resistance levels against a pathogen susceptible to the protein according to the invention, such as

Phytophthora infestans. Alternatively, the selected plants may be subjected to another round of transformation, for instance to introduce further genes, such as genes encoding chitinases, glucanases, osmotins, magainins or the like, in order to enhance resistance levels, or broaden the resistance to other fungi found not to be susceptible to the protein according to the invention in an *in vitro* assay as described herein.

Other evaluations may include the testing of fungal resistance under field conditions, checking fertility, yield, and other characteristics. Such testing is now routinely performed by persons having ordinary skill in the art.

Following such evaluations, the transformed plants may be grown directly, but usually they may be used as parental lines in the breeding of new varieties or in the creation of hybrids and the like.

Many plant proteins exhibit antifungal effects, some however do not do so as such, but yield a significant synergistic antifungal effect if used in combination with other plant proteins. In European Patent Application 440 304 A1 it was disclosed that simultaneous relative over-expression of a plant expressible glucanase gene in conjunction with a basic chitinase from tobacco in transgenic plants results in a higher level of resistance to fungi than in plants expressing a plant expressible class-I chitinase alone.

Both chitinases, glucanases, osmotins, magainins and the new antifungal protein according to the invention accumulate in infected plant tissues upon an incompatible pathogen-plant interaction. From this observation and the fact that several proteins are found to synergise each others antifungal effects, we envision, that the antifungal protein according to the invention may be suitably used in conjunction with other proteins that are associated with pathogen resistance.

Examples of proteins that may be used in combination with the proteins according to the invention include, but are not limited to, ß-1,3-glucanases and chitinases which are obtainable from barley (Swegle M. *et al.*, 1989, Plant Mol. Biol. 12, 403-412; Balance G.M. *et al.*, 1976, Can. J. Plant Sci. 56, 459-466 ; Hoj P.B. *et al.*, 1988, FEBS Lett. 230, 67-71; Hoj P.B. *et al.*, 1989, Plant Mol. Biol. 13, 31-42 1989), bean (Boller T. *et al.*, 1983, Planta 157, 22-31; Broglie K.E. *et al.* 1986, Proc. Natl. Acad. Sci. USA 83, 6820-6824; Vögeli U.

et al., 1988 Planta 174, 364-372); Mauch F. & Staehelin L.A., 1989, Plant Cell 1, 447-457); cucumber (Motraux J.P. & Boller T. (1986), Physiol. Mol. Plant Pathol. 28, 161-169); leek (Spanu P. *et al.*, 1989, Planta 177, 447-455); maize (Nasser W. *et al.*, 1988, Plant Mol. Biol. 11, 529-538), oat (Fink W. *et al.*, 1988, Plant Physiol. 88, 270-275), pea (Mauch F. *et al.* 1984, Plant Physiol. 76, 607-611; Mauch F. *et al.*, 1988, Plant Physiol. 87, 325-333), poplar (Parsons, T.J. *et al*, 1989, Proc. Natl. Acad. Sci. USA 86, 7895-7899), potato (Gaynor J.J. 1988, Nucl. Acids Res. 16, 5210; Kombrink E. *et al.* 1988, Proc. Natl. Acad. Sci. USA 85, 782-786; Laflamme D. and Roxby R., 1989, Plant Mol. Biol. 13, 249-250), tobacco (*e.g.* Legrand M. *et al.* 1987, Proc. Natl. Acad. Sci. USA 84, 6750-6754; Shinshi H. *et al.* 1987, Proc. Natl. Acad. Sci. USA 84, 89-93), tomato (Joosten M.H.A. & De Wit P.J.G.M. 1989, Plant Physiol. 89, 945-951), wheat (Molano J. *et al.*, 1979, J. Biol. Chem. 254, 4901-4907), and the like.

To obtain transgenic plants capable of constitutively expressing more than one chimeric gene, a number of alternatives are available including the following:

A. The use of DNA, *e.g* a T-DNA on a binary plasmid, with a number of modified genes physically coupled to a selectable marker gene. The advantage of this method is that the chimeric genes are physically coupled and therefore migrate as a single Mendelian locus.

B. Cross-pollination of transgenic plants each already capable of expressing one or more chimeric genes, preferably coupled to a selectable marker gene, with pollen from a transgenic plant which contains one or more chimeric genes coupled to another selectable marker. Afterwards the seed, which is obtained by this crossing, maybe selected on the basis of the presence of the two selectable markers, or on the basis of the presence of the chimeric genes themselves. The plants obtained from the selected seeds can afterwards be used for further crossing. In principle the chimeric genes are not on a single locus and the genes may therefore segregate as independent loci.

C. The use of a number of a plurality chimeric DNA molecules, *e.g.* plasmids, each having one or more chimeric genes and a selectable marker. If the frequency of co-transformation is high, then selection on the basis of only one marker is sufficient. In other cases, the selection on the basis of more than one marker is preferred.

D. Consecutive transformation of transgenic plants already containing

a first, second, (etc), chimeric gene with new chimeric DNA, optionally comprising a selectable marker gene. As in method B, the chimeric genes are in principle not on a single locus and the chimeric genes may therefore segregate as independent loci.

E. Combinations of the above mentioned strategies.

The actual strategy may depend on several considerations as maybe easily determined such as the purpose of the parental lines (direct growing, use in a breeding programme, use to produce hybrids) but is not critical with respect to the described invention.

In this context it should be emphasised that plants already containing chimeric DNA capable of encoding antifungal proteins may form a suitable genetic background for introducing chimeric DNA according to the invention, for instance in order to enhance resistance levels, or broaden the resistance. The cloning of other genes corresponding to proteins that can suitably be used in combination with DNA, and the obtention of transgenic plants, capable of relatively over-expressing same, as well as the assessment of their effect on pathogen resistance *in planta* , is now within the scope of the ordinary skilled person in the art.

The obtention of transgenic plants capable of expressing, or relatively over-expressing, proteins according to the invention is a preferred method for counteracting the damages caused by fungi, such as Oomycetes like *Phytophthora infestans*, as will be clear from the above description. However, the invention is not limited thereto. The invention clearly envisions also the use of the proteins according to the invention as such, preferably in the form of a fungicidal composition. Fungicidal composition include those in which the protein is formulated as such, but also in the form of host cells, such as bacterial cells, capable of producing the protein thereby causing the pathogen to be contacted with the protein. Suitable host cells may for instance be selected from harmless bacteria and fungi, preferably those that are capable of colonising roots and/or leaves of plants. Example of bacterial hosts that may be used in a method according to the invention are strains of *Agrobacterium, Arthrobacter, Azospyrillum, Pseudomonas, Rhizobacterium,* and the like, optionally after having been made suitable for that purpose.

Compositions containing antifungal proteins according to the invention may comprise in addition thereto, osmotin-like proteins as

defined in (WO91/18984). Independently, the invention provides antifungal compositions which further comprise inhibitory agents such as classical fungal antibiotics, SAFPs and chemical fungicides such as polyoxines, nikkomycines, carboxymides, aromatic carbohydrates, carboxines, morpholines, inhibitors of sterol biosynthesis, organophosphorus compounds, enzymes such as glucanases, chitinases, lysozymes and the like. Either per se, or in combination with other active constituents, the antifungal protein of the invention should be applied in concentrations between 1 ng/ml and 1 mg/ml, preferably between 2 ng/ml and 0.1 mg/ml, within pH boundaries of 3.0 and 9.0. In general it is desired to use buffered preparations, *e.g.* phosphate buffers between 1mM and 1M, preferably between 10 mM and 100mM, in particular between 15 and 50 mM, whereby in case of low buffer concentrations it is desired to add a salt to increase ionic strength, preferably NaCl in concentrations between 1 mM and 1M, preferably 10 mM and 100 mM.

Plants, or parts thereof, which relatively over-express a protein according to the invention, including plant varieties, with improved resistance against fungal diseases, especially diseases caused by Oomycetes like *Phytophthora* and *Pythium* may be grown in the field, in the greenhouse, or at home or elsewhere. Plants or edible parts thereof may be used for animal feed or human consumption, or may be processed for food, feed or other purposes in any form of agriculture or industry. Agriculture shall mean to include horticulture, arboriculture, flower culture, and the like. Industries which may benefit from plant material according to the invention include but are not limited to the pharmaceutical industry, the paper and pulp manufacturing industry, sugar manufacturing industry, feed and food industry, enzyme manufacturers and the like.

The advantages of the plants, or parts thereof, according to the invention are the decreased need for fungicide treatment, thus lowering costs of material, labour, and environmental pollution, or prolonging shelf-life of products (e.g. fruit, seed, and the like) of such plants. Plants for the purpose of this invention shall mean multicellular organisms capable of photosynthesis, and subject to some form of fungal disease. They shall at least include angiosperms as well as gymnosperms, monocotyledonous as well as dicotyledonous plants.

The phrase "plants which relatively over-express a protein" shall mean plants which contain cells expressing a transgene-encoded protein which is either not naturally present in said plant, or if it is present by virtue of an endogenous gene encoding an identical protein, not in the same quantity, or not in the same cells, compartments of cells, tissues or organs of the plant. It is known for instance that proteins which normally accumulate intracellularly may be targeted to the apoplastic space.

According to another aspect of the invention the regulatory region of a plant gene coding for the antifungal protein of the invention may be used to express other heterologous sequences under the control thereof. The use of a regulatory element of at least 1000 bp directly upstream of the gene coding region is sufficient for obtaining expression of any heterologous sequence.

Heterologous sequences in this respect means gene regions not naturally associated to said regulatory region, and they comprise both different gene coding regions, as well as antisense gene regions. Heterologous coding sequences that may be advantageously expressed in the vascular tissue comprise those coding for antipathogenic proteins, *e.g.* insecticidal, bactericidal, fungicidal, and nematicidal proteins. In such a strategy it may prove exceptionally advantageous to select a protein with activity against a pathogen or pest which has a preference for phloem as source of nutrients (*e.g.* aphids), or as entrance to invade the plant. Examples are extensin, lectin, or lipoxidase against aphids (See WO93/04177). Assuming that the regulatory region according to the invention is active in xylem, chitinases and glucanases may be expressed under the control of said regulatory region to combat *Fusarium, Verticillium, Alternaria* and *Ceratocystus* species.

The use of the regulatory region according to the invention may also be used advantageously to regulate or control phloem transport processes. Numerous other applications will readily occur to those of skill in the art.

The expression of part of (part of) an endogenous gene in the antisense orientation (such as disclosed in EP 0 233 399 A), can effectively down-regulate expression of said endogenous gene, with interesting applications. Moreover, the gene encoding the antifungal protein according to the invention itself may be down-regulated using

the antisense approach which may help establishing the nature and function of the protein. The regions responsible for tissue-specific expression may be unravelled further using the GUS-marker in a way analogous to the way illustrated herein.

The following state of the art may be taken into consideration, especially as illustrating the general level of skill in the art to which this invention pertains.
EP-A 392 225 A2; EP-A 440 304 A1; EP-A 460 753 A2; WO90/07001 A1; US Patent 4,940,840.

Evaluation of transgenic plants

Subsequently transformed plants are evaluated for the presence of the desired properties and/or the extent to which the desired properties are expressed. A first evaluation may include the level of expression of the newly introduced genes, the level of fungal resistance of the transformed plants, stable heritability of the desired properties, field trials and the like.

Secondly, if desirable, the transformed plants can be crossbred with other varieties, for instance varieties of higher commercial value or varieties in which other desired characteristics have already been introduced, or used for the creation of hybrid seeds, or be subject to another round of transformation and the like.

Synergy

The combination of one of the antifungal protein according to the instant invention and other antifungal proteins of plant or microbial source are predicted to show a drastic synergistic antifungal effect. Similar synergistic antifungal effects were shown if combinations of antifungal CBPs or Chi-V are combined with either ß-1,3-glucanases or chitinases from other plant origins. Apparently, the synergizing effect of combinations of pathogen induced proteins is a more general phenomenon that has important consequences for the engineering of fungal resistant plants.

Plants, or parts thereof of commercial interest, with improved resistance against phytopathogenic fungi can be grown in the field or in greenhouses, and subsequently be used for animal feed, direct consumption by humans, for prolonged storage, used in food- or other

industrial processing, and the like. The advantages of the plants, or parts thereof, according to the invention are the decreased need for fungicide treatment, thus lowering costs of material, labour, and environmental pollution, or prolonged shelf-life of products (e.g. fruit, seed, and the like) of such plants.

EXPERIMENTAL PART

Standard methods for the isolation, manipulation and amplification of DNA, as well as suitable vectors for replication of recombinant DNA, suitable bacterium strains, selection markers, media and the like are described for instance in Maniatis *et al.*, molecular cloning: A Laboratory Manual 2nd. edition (1989) Cold Spring Harbor Laboratory Press; DNA Cloning: Volumes I and II (D.N. Glover ed. 1985); and in: From Genes To Clones (E.-L. Winnacker ed. 1987).

In vitro antifungal assay on Oomycetes

The antifungal activity was monitored during purification in a microtiter plate assay using the fungi *Phytophthora infestans* and *Pythium ultimum* according to Woloshuk *et al.*, 1991. In each well of a 24-well microtiter dish 250 µl potato dextrose agar (PDA) was pipetted. Fungal spores in the case of *Phytophthora infestans* and hyphal fragments in the case of *Pythium ultimum* were suspended in water and 400-600 spores or 200 fragments in 50 µl were added to the wells. Subsequently 100 µl filter sterilized (0.22 µm filter) protein solution (in 50 mM MES, pH 6.0) was added. Microtiter dishes were wrapped with Parafilm and incubated at room temperature. At several timepoints after the initiation of incubation the fungus was monitored microscopically for effects of the added protein. After 2-3 days the mycelium of the growing fungus in the wells was stained with lactophenol cotton blue and the extent of growth was estimated. GI: growth inhibition; a scale of 0 - 4 is used, 0 = no visible inhibition, 1 = weak inhibition (0 to 30%) inhibition, 2 = moderate (30 to 60%) inhibition, 3 = strong (60 to 90%) inhibition, 4 = very strong (100%) inhibition.

EXAMPLE 1

Purification of an antifungal protein MS59 from sunflower induced with salicylic acid

Leaves of 7 to 8 weeks old sunflower (*Helianthus annuus* cv. zebulon) plants were sprayed daily for 5 times with 10 mM sodium salicylate. After 3 hours the plants were extensively rinsed with water to remove the sodium salicylate. Three days after the final spray, leaves (400 gram) were harvested into liquid nitrogen and homogenized at 4°C in 500 ml 0.5 M NaOAc pH5.2, and 4 gram active carbon, using a Waring blendor. The homogenate was filtered over four layers of cheese cloth and subsequently the filtrate was centrifuged for 50 minutes at 20,000 g at 4°C and desalted by passage through a Sephadex G25 column (medium course; Pharmacia), length 60 cm, diameter 11.5 cm, equilibrated in 40 mM NaOAc pH5.2. The desalted protein solution was stored overnight at 4°C and subsequently centrifuged for 45 minutes at 20,000 g at 4°C. The supernatant was passed through a S-sephadex (Fast-flow, Pharmacia) column, length 5 cm, diameter 5 cm, which was equilibrated with 40 mM NaOAc pH 5.2. The column was washed with the above mentioned buffer (flow rate 400 to 500 ml/hr) until the OD_{280} dropped to zero. The bound proteins were eluted using a 400 mM NaCl in 200 ml of the above mentioned buffer.

After dialysis against 50 mM MES pH 6.0 the eluate was analyzed for antifungal activity. Antifungal activity was monitored in a microtiter plate assay using the fungus *Phytophthora infestans* and *Pythium ultimum*. See above for details concerning *in vitro* assaying. Subsequently, cationexchange chromatography was reapplied whereby the eluate was passed through an FPLC Mono-S HR 5/5 (Pharmacia) and eluted with a linear gradient from 0 to 400 mM NaCl. All fractions were analyzed by electrophoresis (Laemmli (1970), Nature 227:680-685) using a 12.5% polyacrylamide gel in the presence of sodium dodecyl sulphate (SDS), using prestained molecular weight markers (15-105 kDa) as reference. Additionally, of all fractions antifungal activity towards *Phytophthora infestans* and *Pythium ultimum* was monitored. Antifungal activity eluted from the column between 45-60 mM NaCl and in all active fractions a 59 kD band was visible. Fractions containing the antifungal activity were pooled and dialysed to 1 M ammonium sulphate in 50 mM potassium phosphate, pH 7. The pool was subjected to hydrophobic interaction chromatography, whereby the sample was applied

to an FPLC Phenyl Superose HR 5/5 (Pharmacia) equilibrated in the same buffer and eluted with a linear decreasing gradient from 1 to 0 M ammonium sulphate in 50 mM potassium phosphate, pH 7. As above again all fractions were analyzed on SDS-PAGE and monitored for antifungal activity. Also the pool of proteins not capable of binding to this column (Flow Through, FT) was thus analyzed at the conditions chosen here. Antifungal was present most abundantly in the FT and secondly also in the fractions eluting between 0.76 and 0.45 mM ammonium sulphate. In both cases a 59 kD protein was visible on SDS-PAGE. FT and the gradient fractions were separately dialysed to 50 mM MES, 0.2 M NaCl and separately chromatographed on a FPLC Superdex 75 HR 10/30 column (Pharmacia) equilibrated to the same buffer. Proteins elute from this column according to their molecular size. In both cases again the presence of a 59 kD protein coincided with antifungal activity towards *Phytophthora infestans* and *Pythium ultimum* as judged from SDS-PAGE and *in vitro* antifungal assays. The 59 kD protein present in the FT of the hydrophobic interaction column was most abundant and termed MS59 and its purification is visualized in Figure 1. Results of its separation over the gelfiltration column and subsequent analysis both on SDS-PAGE and on *Phytophthora infestans* is shown in Figure 2. Several characteristics (antifungal activity, chromatographical properties, molecular mass) of the gradient protein and MS59 indicate that the two proteins are very similar.

To characterize MS59 further its amino acid sequence was partially determined. Therefore, MS59 was separated in the presence of 0.1 mM thioglycolate in the upper reservoir buffer and SDS on a 12.5% polyacrylamide gel, which was prerun for 2 hours at 50 V with 0.05 mM glutathion in the upper reservoir buffer. The gel was stained with 5% (w/v) Serva Blue G in 45% (v/v) methanol and 10% acetic acid for 30 minutes and destained in 20% (v/v) acetic acid for 30 minutes and the 59 kDa band was cut out and sequenced using Edman degradation on an Applied Biosystems 477A protein sequencer according to the protocol provided by the manufacturer. N-terminal amino acid sequencing of MS59 revealed that the N-terminus was blocked. To obtain internal sequences, MS59 was digested with trypsin. Trypsin cleaves protein at arginine and lysine residues. The digestion products were separated on a reversed-phase column and analyzed by Edman degradation. Two tryptic fragments were sequenced: Pep1 and Pep2. Of Pep1 25 amino acid

residues were identified: S-I-N-V-D-I-E-Q-E-T-A-W-V-Q-A-G-A-T-L-G-E-V-Y-Y-R (SEQ ID NO: 1).

The amino acid sequence is given using the one-letter code. Of Pep2 a further 25 amino acid residues were identified: D-P-S-F-P-I-T-G-E-V-Y-T-P-G-(?)-S-S-F-P-T-V-L-Q-N-Y (SEQ ID NO: 2).

The amino acid residue between brackets could not be identified unambiguously.

Polyclonal antibodies were raised against MS59. Firstly, the whole MS59 protein was used to raise antibodies. Therefore, MS59 was separated on SDS-polyacrylamide gels and the MS59 band was cut out after visualization. The band (approximately 100 µg protein) was dehydrated in EtOH, dialysed extensively against phosphate buffered saline (PBS), ground in a mortar and injected into a rabbit. After one month the rabbit was boostered every 2 weeks with 50 µg MS59 protein. After 15 weeks the rabbit was sacrificed. The antiserum was further purified on a horse radish peroxidase (HRP) column, since the antiserum was cross-reactive with a lot of probably glycosylated proteins. After purification the antiserum specifically recognized a protein band of 59 kDa and showed some minor cross reaction with a low number of other proteins.

In a second approach peptides of MS59, which were synthesized according to the amino acid sequence of SEQ ID NO:1 and SEQ ID NO:2, or parts thereof, coupled to bovine serum albumine were used to raise antiserum. This antiserum, which did not need a purification on a HRP column, specifically recognized a protein band of 59 kDa.

EXAMPLE 2

Elution of antifungal protein from native PAGE and subsequent testing

It is obvious from Figure 1 that MS59 is not completely pure. To further ensure that indeed the 59 kDa protein is responsible for the observed antifungal activity, the fraction containing the peak amount of 59 kDa was electrophoresed on a native gel, using the same system as described above however without SDS and without boiling the samples before loading. The gel lane was sliced in 0.5 cm horizontal pieces and each piece was eluted individually for 48 hours in 50 mM Mes, pH 6. After centrifugation the resulting supernatant was analyzed both on SDS-PAGE and *in vitro* for antifungal activity. Results are shown in

Figure 3. Only in those fractions containing MS59, was antifungal activity observed against *Phytophthora infestans* and *Pythium ultimum*.

EXAMPLE 3

***In vitro* antifungal assays on non-Oocymetes**

All fungi were cultured on potato dextrose agar (Difco) at 25°C, except *Botrytis cinerea* and *Phoma lingam* which were grown on oat meal agar (Difco) at 25°C. *Phytophthora infestans* was grown on rye agar at 18°C in the dark (Caten and Jinks, 1968). *Botrytis cinerea* and *Phoma lingam* were cultivated under UV. Spores of sporulating fungi were harvested by flooding the agar plates with water. The spore concentration was adjusted to 10,000 sp/mL. In the case of *Rhizoctonia solani* and *Pythium ultimum* liquid shake cultures were grown in potato dextrose broth at 25°C. To prepare inoculum from these shake cultures, mycelium was harvested and vortexed for 1 minute. After passage through a fine sieve, inoculum density was adjusted to 2500 - 5000 fragments, of 1 to 3 cells each, per mL.

In case of sporulating fungi, all were tested both with and without pregerminating the spores before application of the protein samples. In case of non-sporulating fungi, hyphal fragments were used. Fractions eluting from the Mono-S, pH 6 were assayed for the presence of antifungal activity. As positive control *Phytophthora infestans* was tested. The peak of MS59 is located in fraction 4. Results are shown in Table 1.

TABLE 1

Antifungal effects of MS59 containing fractions from Mono-S, pH 6

fungus	spore stage *)	fraction number 1	2	3	4	5	6	78
Botrytis cinerea	spore	0	0	0	0	0	0	0	0
	germl.	0	0	0	0	0	0	0	0
Fusarium oxysporum	spore	0	0	0	0	0	0	0	0
	germl.	2	2	2	2	3	3.5	3.5	3.5
Fusarium solani	spore	0	0	0	0	0	0	0	0
Alternaria solani	spore	0	0	0	0	0	0	0	0

	germl.	0	0	0	0	0	0	0	0
Phoma lingam	spore	0	0	0	0	0	0	0	0
	germl.	0	0	0	0	0	0	0	0
Colletotrichum coccodes	spore	0	0	0	0	0	0	0	0
	germl.	0	0	0	0	0	0	0	0
Rhizoctonia solani	hyph.	0	0	0	0	0	0	0	0
Phytophthora infestans	spore	0	2	2	4	3.5	2	1	0
Phytophthora nicotianae	hyph	0	1	2	4	4	2	1	0
Phytophthora cactorum	hyph	0	0	2	4	4	1	1	0
Pythium ultimum	hyph	0	0	0	4	4	0	0	0
Pythium sylvaticum	hyph	0	0	0	2	1	0	0	0
Pythium paroecandrum	hyph	0	0	0	2	2	0	0	0

*) spore = no pregermination, germl = germination until the germtube is 3-5 times the length of the spore, hyph. = hyphal fragments were used as starting inoculum.

GI: growth inhibition; a scale of 0 - 4 is used, 0 = no visible growth inhibition, 1 = weak (0 to 30%) inhibition, 2 = moderate (30 to 60%) inhibition, 3 = strong (60 to 90%) inhibition, 4 = very strong (100%) inhibition.

As can be seen *Phytophthora* and *Pythium* spp., appeared very sensitive to MS59, indicating the specificity of MS59 for Oomycetes.

EXAMPLE 4

Identification and characterization of genes homologous to the deduced MS59 nucleotide sequence

Based on the amino acid sequences of pep1 (a.a. 12 to 22 of SEQ ID NO: 1) and pep2 (a.a. 2 to 12 of SEQ ID NO: 2), primers were designated for PCR. Genomic DNA was isolated from sunflower cv. Zebulon and PCR primers 4 (5'AAC TTC TCC IAG IGT IGC ICC IGC TTG IAC CCA3', SEQ ID NO: 3) and 5 (5'GAT CCI TCT TTC CCI ATT ACT GGI GAG GTT TA3', SEQ ID NO: 4) were used to amplify a 354 bp DNA fragment from the sunflower genome with PCR. PCR products corresponding to this fragment size were cloned (SEQ ID NO: 5). Sequence analysis of the product revealed the

presence of an uninterrupted Open Reading Frame (ORF) (SEQ ID NO: 6) of which the first and last stretch of amino acids corresponded with the amino acid sequences of SEQ ID NO: 1 and SEQ ID NO: 2. Several clones sequenced contained point mutations, varying from 1 to 4 in this PCR fragment. All but one of these mutations were silent mutations (nucleotide nr 57 T to C, nucleotide nr 63 C to A, nucleotide nr 225 A to G) which therefore did not alter amino acid sequences encoded. One clone however did contain a point mutation (nucleotide nr 203 G to A) which altered the amino acid sequence at amino acid 68 from Arg to Lys.

A southern blot of sunflower genomic DNA, probed with SEQ ID NO: 5 indicated the existence of multiple homologous sequences in the genome. Using SphI, 6 bands were detected, EcoRV 5 bands, SpeI 3 bands and NdeI 4 bands. With other enzymes 3-4 bands were previously discerned. This analysis suggests the existence of 3 genes with (partial) homology to the ms59 sequences.

New PCR primers were developed based on the non-variable areas between the original PCR primer sequences. Primers: for 3' RACE: 5' CAG GCA GCT GTG GTT TGT GGC 3' (SEQ ID NO: 7),
for 5' RACE: 5' GTC CAC AAT GAA GAA GGG TTG 3' (SEQ ID NO: 8) and for nested 3'RACE: 5' ACG TAG ATA TCG AAC AAG AAA CCG C 3' (SEQ ID NO: 9).

Poly(A) containing RNA was isolated from sunflower leaf material that was induced by spraying 5 times with a 10 mM sodium salicylate solution. cDNA was prepared and 5' and 3' RACE PCR reactions were performed as described in the instructions of the Marathon™ kit (Clontech laboratories, Inc., Palo Alto, CA). Partial cDNA clones were isolated by 5' and 3' RACE PCR reactions. Sequence analysis confirmed the identity of the partial cDNA clones.
Again new PCR and nested PCR primers were developed based on newly obtained sequence information from cloned 5' and 3' RACE PCR products. Primers for 5'RACE: 5' CTG GGG AAG CCC GTG TAG TAA AGC 3' (SEQ ID NO:11), 5' CGG GAA GTT GCA GAA GAT TGG GTT G 3' (SEQ ID NO:13), for nested 5'RACE: 5' GAG CAA GAG AAG AAG GAG AC 3' (SEQ ID NO:14), for 3' RACE: 5' GCT TTA CTA CAC GGG CTT CCC CAG 3' (SEQ ID NO: 10), and for nested 3' RACE: 5' GGT ACT CCA ACC ACG GCG CTC 3' (SEQ ID NO:12).

Four partial cDNA clones were isolated which together encode all of the Open Reading Frame including a putative signal peptide followed by an approximately 59 kDa protein, and 5' and 3' UTR's (untranslated regions)(SEQ ID NO: 15). A full length cDNA clone of 1784 bp, of which the ORF (pos. 21 to pos. 1608) encodes 529 amino acid residues (SEQ ID NO:16), could be assembled out of these four partial cDNA clones and the PCR fragment mentioned above (SEQ ID NO: 5).
The amino-terminal signal sequence (Von Heijne *et al.*, 1983 and Von Heijne, 1985) is not likely represented by the first 19 amino acid residues. A prediction of the putative cleavage site was made.

The amino acid sequence of this cDNA clone was used in a blast homology search. This sequence revealed high homology to the Berberine Bridge Enzymes (BBE) from Californian poppy (*Eschscholtzia californica*) (Dittrich and Kutchan, 1991, Proc. Natl. Acad. Sci. USA 88, 9969-9973) and Opium poppy (*Papaver somniferum*) (Facchini *et al.*, 1996, Plant Physiol. 112, 1669-1677).
Blast screening of Expressed Sequence Tag (dbEST) databases with the amino acid sequence as shown in SEQ ID NO: 16 revealed homologues of the MS59 protein in *Arabidopsis thaliana* (SEQ ID NO: 21 to SEQ ID NO: 46) and rice (SEQ ID NO: 47).
The EST sequences are listed in the sense orientation considering the orientation of homology to MS59. Sequences of the EST clones were altered by inserting one or two extra unknown nucleotides (N or NN) at frameshift positions in order to obtain one single translation frame with homology to MS59.

TABLE 2

The names and presumed frameshifts of the different EST's in which homology with the MS59 a.a. sequence is optimal. In the columns with frameshift 1 and frameshift 2 the position of the frameshift and the shift (frame--->frame) are listed. The (-) mark means, no frameshift present.

SEQ ID NO:	EST name	GenBank accession	Frame (1, 2, 3)	Frameshift	Frameshift
19	ATTS5925	F19886	2	-	-
20	ATTS0345	Z17771	2	202, 2 ---> 1	-
21	ATTS5268	F14356	2	298, 2 ---> 3	-
22	TC13883	-	1	177, 1 ---> 2	-
23	TC11550	-	2	-	-
24	P_16053	R84094	1	-	-
25	P_22214	N97049	2	310, 2 ---> 3	-
26	P_16873	R90518	3	317, 3 ---> 2	-
27	ATTS2532	Z30784	3	188, 3 ---> 1	312, 1 ---> 2
28	TC11456	-	3	-	-
29	P_8818	T45555	1	98, 1 ---> 3	-
30	P_21340	N96011	1	-	-
31	P_22585	W43206	2	367, 2 ---> 1	-
32	Q_ATTS2533	Z30785	2	-	-
33	P_17333	H76902	2	-	-
34	P_9615	T46352	1	-	-
35	Q_ATTS2959	Z33920	2	-	-
36	P_2730	T20722	1	-	-
37	TC9870	-	2	-	-
38	P_14876	H36354	2	241, 2 ---> 1	-
39	P_21353	N96040	1	89, 1 ---> 2	-
40	Q_ATTS3343	Z34583	1	-	-
41	Q_ATTS4954	F14032	2	139, 2 ---> 1	-
42	Q_ATTS1606	Z26512	2	-	-
43	P_7866	T44603	1	222, 1 ---> 3	-
44	AA0410042	24308	2	421, 2 ---> 1	-
45	RICS2381A	D40415	3	-	-

EXAMPLE 5

Tailoring a MS59 clone for expression in transgenic plants

PCR primers were developed based on the sequence around the ATG start codon and the TGA stop codon for cloning of the open reading frame (ORF). A NcoI restriction site was introduced at the ATG start codon for fusion to a constitutive promoter by PCR using primer: 5' CC GCC ATG GAG ACT TCC ATT CTT ACT C 3' (SEQ ID NO:16). The second codon of the ORF was changed from caa (Q) to gag (E) as a result of the introduced NcoI restriction site.

Downstream of the TGA stop codon a BamHI restriction site was introduced by PCR using primer: 5' GCC GGA TCC TCA AGA TGA CAA AGT TGG GAT GCT 3' (SEQ ID NO:18).

Using a PCR reaction with *Pfu* DNA polymerase, we amplified the entire ORF, using the PCR primers to introduce the NcoI restriction site on the startcodon ATG and the BamHI restriction enzyme recognition site just downstream of the stopcodon. The integrity of the DNA sequence was confirmed by sequencing (SEQ ID NO:19). The entire ORF was linked to a constitutive promoter which allows high level protein expression in most parts of the plant. After the ORF a 3' untranslated region of the potato proteinase inhibitor II (Thornburg *et al.*, 1987, Proc. Natl. Acad. Sci. USA 84, 744-748), which contains sequences needed for polyadenylation (An *et al.*, 1989, Plant Cell 1, 115-122), was introduced. The chimeric gene produced was introduced into binary vector pMOG800 (deposited at the Centraal Bureau voor Schimmelcultures, Baarn, The Netherlands, under CBS 414.93, on August 12, 1993). The resulting clones pMOG1144, which harbours the MS59 construct under control of the 35S promoter and pMOG1180, which harbours the MS59 construct under control of the ocs-mas hybrid promoter (WO95/14098) were introduced in *Agrobacterium tumefaciens* strain EHA105, suitable for transformation of target crops tomato and potato, strain MOG101 for transformation of tobacco and *Arabidopsis* and MOG301 for transformation of *Brassica napus*.

In addition to this single gene construct, two 4-gene constructs are made designated;

a) pMOG1145, which contains the tobacco genes encoding class I ß-1,3-glucanase, class I chitinase and AP24 under control of heterologous promoters and the MS59 gene construct as described before, and

b) pMOG1146, which contains the tobacco genes encoding class I ß-1,3-

glucanase, class I chitinase and 16 kDa basic PR-1 under control of heterologous promoters and the described MS59 gene construct. These four gene constructs are constructed in the binary vector pMOG800 and introduced into *Agrobacterium tumefaciens* strain EHA105 for plant transformation.

EXAMPLE 6

Production and analysis of transgenic tobacco, tomato, potato, carrot, and Brassica napus plants containing the MS59 gene construct

Using *Agrobacterium* mediated transformation system binary constructs containing the MS59 gene construct as described in Example 5 are introduced into different crop species including, tobacco, tomato, potato, carrot, and *Brassica napus*.

The transgenic shoots of these different plant species are regenerated into whole plants and subsequently, primary transformants are analyzed for expression of the newly introduced MS59 gene. For this analysis use is made of Western blotting techniques, using antibodies against MS59 specific peptide coupled to BSA. All antisera are diluted 1:5,000. A concentration series of purified proteins (12.5, 25, 50 and 100 ng) is used to judge the expression level of the introduced proteins in the transgenic plants. Transgenic samples are homogenized in 50 mM sodium acetate buffer pH 5.2 and the extracts are clarified by centrifugation. The supernatants are either directly analyzed or left overnight to precipitate on ice. Overnight precipitation is always followed by clarification (by centrifugation). The protein concentration of the supernatants obtained in either way is determined using Bradford reagent (Bradford 1976, Anal. Biochem. 72, 248-254) and BSA as the standard protein. As much protein as possible (but never more than 10 µg) is loaded on a 12,5 % SDS-PAA gel (Laemmli, supra) and immunoblotted as previously described (Ponstein *et al.* supra).

Extracts from leaves of ms59-transgenic tobacco, tomato, potato, carrot, and *Brassica napus* plants are made by pottering leaf fragments in a buffer containing 50 mM NaAc, (pH = 5.2). After this, insoluble protein is removed by centrifugation. Total soluble protein content is measured and the equivalent of 10 µg was loaded on a SDS-gel. After running the gel the proteins were transferred to blot. This blot is

developed using the antiserum raised against purified MS59. The MS59-specific antiserum is used in a 1:5,000 dilution. Purified MS59 is also run alongside on the gel, and is included for reference.

A number of transformed plants will be selected based on their high level expression of MS59 protein and S1 progeny plants are tested in fungal infection assays.

EXAMPLE 7

Expression of MS59 in *E. coli* and preparation of antiserum

A PCR fragment containing the presumed mature portion of MS59 was introduced in vector pET30c (Novagen, Madison, WI), and the correct insertion of the fragment is confirmed using DNA sequencing. Then, the plasmid was introduced into *E. coli* BL21 (DE3) pLysS (Novagen, Madison, WI). Small scale cultures (2 ml) of several colonies were then started of which half is induced by the addition of IPTG to 1mM final concentration. Total extracts from *E.coli* were run on SDS gels and analyzed by Coomassie Brilliant Blue staining.

Several clones exhibited strong overexpression of the MS59 protein. A clone which had strong overexpression was selected for a large scale culture. Five hundred ml of LB supplemented with 0.4 mM glucose was inoculated with a culture of this *E. coli* and grown to an optical density of 0.5-0.7. Then, IPTG was added to a final concentration of 1 mM and protein production allowed for 3 hours at 30°C. A large proportion of the MS59 protein was found in the insoluble protein fraction, a small amount appeared soluble. The resulting insoluble protein preparation contained mainly MS59 protein. This preparation is used for raising antibodies. Two rabbits (NZW) are injected at regular intervals with MS59 protein. Antiserum from these rabbits is tested on Western blots containing purified *E.coli*- and sunflower derived MS59. Both proteins will be readily detected by the antiserum using standard Western blot procedures.

EXAMPLE 8

In vitro* antifungal assay of *E.coli*-produced MS59 on *Phytophthora* and *Pythium

The MS59 protein produced in *E. coli* contained N-terminal trxA -, His-and a S-Tags. The His-tag was used for purification of the soluble MS59 on an IMAC (immobilized metal affinity chromatography) column, charged with Ni^{2+}. Bound protein was eluted by increasing the imidazole concentration. The peak fraction from this purification contains some contaminating *E. coli* proteins.

The peak fraction of this MS59 purification was dialysed into 50 mM MES, pH 6.0, and used in an *in vitro* assay with *Phytophthora infestans* and *Pythium ultimum*. For the standard setup of the in vitro antifungal assay with *Phytophthora infestans* and *Pythium ultimum* see above. As control treatment we assayed an unrelated His-tagged protein purified from the same expression host, with some *E.coli protein* background. Also a boiled MS59 control (heated 10 minutes at 100°C) was included. Approximately 40 ng of fusion protein was tested in the *Phytophthora infestans* assay, twice that amount was used for the *Pythium ultimum* inhibition assay.

Microtiter dishes were wrapped with Parafilm and incubated in the dark at room temperature. After 2-3 days the mycelium of the growing fungus in the wells was stained with lactophenol cotton blue and the extent of growth was estimated.

TABLE 3

Antifungal effects of MS59 from *E.coli* on *Phytophthora infestans*

Fraction	$MS59_{E. coli}$	$MS59_{E. coli}$ boiled	His-protein E.coli	MES buffer
Growth inhibition	4	0	0	0
amount of extract	5µl	5µl	5µl	

Growth inhibition (GI) is scored visually on a linear scale of 0 (no inhibition) to 4 (complete growth inhibition).

TABLE 4

Antifungal effects of MS59 from *E.coli* on *Pythium ultimum*

Fraction	$MS59_{E. coli}$	$MS59_{E. coli}$ boiled	His-protein E.coli	MES buffer
Growth inhibition	4	0	0	0
amount of extract	10µl	10µl	10µl	

Microscopical analysis of the wells indicate the rapid germination and subsequent growth of *Phytophthora infestans* zoospores in each of the controls. Germination is near completely inhibited in the reactions containing the MS59 protein from *E. coli*. Some spores do germinate, but hyphal tip growth appears to stop soon after initiation. After 48 hours growth of Phytophthora infestans mycelium is abundant in the controls, but almost undetectable in the assay containing MS59. Even after 72 hours no substantial growth is observed. Fungal hyphae appear somewhat granular and thickened in the reactions containing MS59 protein. Examples of the characteristic patterns of fungal growth in

incubations with and without *E. coli*-produced MS59 are depicted in figure 4. After 48 and 72 hours fungal growth in the control incubations is so extensive no photographic material could be gathered. Incubations in the presence of MS59 leads to complete blockage of further growth, the germination tubes observed at 24 hours do not noticeably extend further.

Likewise, in the *Pythium ultimum* inhibition assay, where mycelium fragments are used, no growth is apparent upon treatment with MS59 (see fig. 5). After 24 hours the control reactions were completely overgrown by mycelium. Only small mycelium fragments are at that stage apparent in the MS59-treated sample.

EXAMPLE 9

Purification of MS59 transproteins from tobacco transgenics

Transgenic tobacco plants were produced expressing MS59 constitutively. Levels of expression are determined using Western analysis. Extracts of the transgenic material are assayed for *in vitro* growth inhibitory activity against *Phytophthora infestans and Pythium ultimum*. The extracts are made by grinding up leaf tissue from transgenic plants in 50 mM NaOAc, pH=5.2. After repeated centrifugation, overnight incubation on ice and an additional centrifugation step, the supernatant is dialysed to 15 mM potassium phosphate + 20 mM sodium chloride, pH =6. After filter sterilisation, 100 µg protein in 100 µl dialysis buffer is added per well containing 250 µl PDA and 50 µl water containing 400-600 spores. Growth inhibition is scored after 3 to 4 days.

EXAMPLE 10

Introduction of the four genes construct containing Chi-I, Glu-I, AP24 and MS59 under control of a constitutive plant promoter, into tomato, potato, carrot, *Brassica napus* and *Arabidopsis*

Using *Agrobacterium* mediated transformation system binary construct pMOG1145 and pMOG1180 containing the genes encoding Chi-I, Glu-I, AP24 and MS59 or pMOG1146 containing the genes encoding Chi-I, Glu-I, bPR-1 and MS59 is introduced into different crop species including, tomato, potato, carrot, *Brassica napus* and Arabidopsis. S1 progeny plants are tested in fungal infection assays.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: MOGEN International nv
(B) STREET: Einsteinweg 97
(C) CITY: Leiden
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): 2333 CB
(G) TELEPHONE: 31-(0)71-5258282
(H) TELEFAX: 31-(0)71-5221471

(ii) TITLE OF INVENTION: Antifungal proteins, DNA coding therefor, and hosts incorporating same.

(iii) NUMBER OF SEQUENCES: 47

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Helianthus annuus
(B) STRAIN: cv. zebulon

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp Val Gln Ala Gly
1               5                  10                  15

Ala Thr Leu Gly Glu Val Tyr Tyr Arg
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Helianthus annuus
(B) STRAIN: cv. zebulon

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr Pro Gly Xaa Ser
1 5 10 15

Ser Phe Pro Thr Val Leu Gln Asn Tyr
20 25

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION: /function= "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AACTTCTCCN AGNGTNGCNC CNGCTTGNAC CCA
33

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION: /function= "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GATCCNTCTT TCCCNATTAC TGGNGAGGTT TA
32

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Helianthus annuus
(B) STRAIN: cv. zebulon

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..354

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GAT CCG TCT TTC CCG ATT ACT GGG GAG GTT TAC ACT CCC GGA AAC TCA
48
Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr Pro Gly Asn Ser
  1               5                  10                  15

TCT TTT CCT ACC GTC TTG CAA AAC TAC ATC CGA AAC CTT CGG TTC AAT
96
Ser Phe Pro Thr Val Leu Gln Asn Tyr Ile Arg Asn Leu Arg Phe Asn
            20                  25                  30

GAA ACT ACC ACA CCA AAA CCC TTT TTA ATC ATC ACA GCC GAA CAT GTT
144
```

```
Glu Thr Thr Thr Pro Lys Pro Phe Leu Ile Ile Thr Ala Glu His Val
          35                  40                  45

TCC CAC ATT CAG GCA GCT GTG GTT TGT GGC AAA CAA AAC CGG TTG CTA
192
Ser His Ile Gln Ala Ala Val Val Cys Gly Lys Gln Asn Arg Leu Leu
     50                  55                  60

CTG AAA ACC AGA AGC GGT GGT CAT GAT TAT GAA GGT CTT TCC TAC CTT
240
Leu Lys Thr Arg Ser Gly Gly His Asp Tyr Glu Gly Leu Ser Tyr Leu
 65                  70                  75                  80

ACA AAC ACA AAC CAA CCC TTC TTC ATT GTG GAC ATG TTC AAT TTA AGG
288
Thr Asn Thr Asn Gln Pro Phe Phe Ile Val Asp Met Phe Asn Leu Arg
                 85                  90                  95

TCC ATA AAC GTA GAT ATC GAA CAA GAA ACC GCA TGG GTC CAA GCC GGC
336
Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp Val Gln Ala Gly
            100                 105                 110

GCC ACC CTC GGA GAA GTT
354
Ala Thr Leu Gly Glu Val
        115
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr Pro Gly Asn Ser
  1               5                  10                  15

Ser Phe Pro Thr Val Leu Gln Asn Tyr Ile Arg Asn Leu Arg Phe Asn
            20                  25                  30

Glu Thr Thr Thr Pro Lys Pro Phe Leu Ile Ile Thr Ala Glu His Val
          35                  40                  45

Ser His Ile Gln Ala Ala Val Val Cys Gly Lys Gln Asn Arg Leu Leu
     50                  55                  60

Leu Lys Thr Arg Ser Gly Gly His Asp Tyr Glu Gly Leu Ser Tyr Leu
 65                  70                  75                  80
```

Thr Asn Thr Asn Gln Pro Phe Phe Ile Val Asp Met Phe Asn Leu Arg
85 90 95

Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp Val Gln Ala Gly
100 105 110

Ala Thr Leu Gly Glu Val
115

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION: /function= "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CAGGCAGCTG TGGTTTGTGG C
21

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION: /function= "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTCCACAATG AAGAAGGGTT G
21

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1
(D) OTHER INFORMATION: /function= "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ACGTAGATAT CGAACAAGAA ACCGC
25

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GCTTTACTAC ACGGGCTTCC CCAG
24

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTGGGGAAGC CCGTGTAGTA AAGC
24

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGTACTCCAA CCACGGCGCT C
21

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CGGGAAGTTG CAGAAGATTG GGTTG
25

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GAGCAAGAGA AGAAGGAGAC
20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1784 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Helianthus annuus
(B) STRAIN: Zebulon

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 21..1611

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATATCACATC TTCTTTCAAC ATG CAA ACT TCC ATT CTT ACT CTC CTT CTT
50
Met Gln Thr Ser Ile Leu Thr Leu Leu Leu
1 5 10

CTC TTG CTC TCA ACC CAA TCT TCT GCA ACT TCC CGT TCC ATT ACA GAT
98
Leu Leu Leu Ser Thr Gln Ser Ser Ala Thr Ser Arg Ser Ile Thr Asp
15 20 25

CGC TTC ATT CAA TGT TTA CAC GAC CGG GCC GAC CCT TCA TTT CCG ATA
146
Arg Phe Ile Gln Cys Leu His Asp Arg Ala Asp Pro Ser Phe Pro Ile
30 35 40

ACC GGA GAG GTT TAC ACT CCC GGA AAC TCA TCT TTT CCT ACC GTC TTG
194

```
    Thr Gly Glu Val Tyr Thr Pro Gly Asn Ser Ser Phe Pro Thr Val Leu
              45                  50                  55

    CAA AAC TAC ATC CGA AAC CTT CGG TTC AAT GAA ACT ACC ACA CCA AAA
 5  242
    Gln Asn Tyr Ile Arg Asn Leu Arg Phe Asn Glu Thr Thr Thr Pro Lys
         60                  65                  70

    CCC TTT TTA ATC ATC ACA GCC GAA CAT GTT TCC CAC ATT CAG GCA GCT
10  290
    Pro Phe Leu Ile Ile Thr Ala Glu His Val Ser His Ile Gln Ala Ala
     75                  80                  85                  90

    GTG GTT TGT GGC AAA CAA AAC CGG TTG CTA CTG AAA ACC AGA AGC GGT
15  338
    Val Val Cys Gly Lys Gln Asn Arg Leu Leu Leu Lys Thr Arg Ser Gly
                   95                 100                 105

    GGT CAT GAT TAT GAA GGT CTT TCC TAC CTT ACA AAC ACA AAC CAA CCC
20  386
    Gly His Asp Tyr Glu Gly Leu Ser Tyr Leu Thr Asn Thr Asn Gln Pro
              110                 115                 120

    TTC TTC ATT GTG GAC ATG TTC AAT TTA AGG TCC ATA AAC GTA GAT ATC
25  434
    Phe Phe Ile Val Asp Met Phe Asn Leu Arg Ser Ile Asn Val Asp Ile
         125                 130                 135

    GAA CAA GAA ACC GCA TGG GTC CAA GCC GGT GCG ACT CTT GGT GAA GTG
30  482
    Glu Gln Glu Thr Ala Trp Val Gln Ala Gly Ala Thr Leu Gly Glu Val
    140                 145                 150

    TAC TAT CGA ATA GCG GAG AAA AGT AAC AAG CAT GGT TTT CCG GCA GGG
35  530
    Tyr Tyr Arg Ile Ala Glu Lys Ser Asn Lys His Gly Phe Pro Ala Gly
    155                 160                 165                 170

    GTT TGT CCA ACG GTT GGC GTT GGT GGG CAT TTT AGT GGT GGT GGG TAT
40  578
    Val Cys Pro Thr Val Gly Val Gly Gly His Phe Ser Gly Gly Gly Tyr
                   175                 180                 185

    GGT AAT TTG ATG AGA AAA TAT GGT TTG TCG GTT GAT AAT ATT GTT GAT
45  626
    Gly Asn Leu Met Arg Lys Tyr Gly Leu Ser Val Asp Asn Ile Val Asp
              190                 195                 200

    GCT CAA ATA ATA GAT GTG AAT GGC AAG CTT TTG GAT CGA AAG AGT ATG
50  674
    Ala Gln Ile Ile Asp Val Asn Gly Lys Leu Leu Asp Arg Lys Ser Met
         205                 210                 215

    GGT GAG GAT TTG TTT TGG GCG ATC ACC GGC GGT GGT GGT GTT AGT TTT
55  722
```

```
Gly Glu Asp Leu Phe Trp Ala Ile Thr Gly Gly Gly Gly Val Ser Phe
    220                 225                 230

GGT GTG GTT CTA GCC TAC AAA ATC AAA CTA GTT CGT GTT CCG GAG GTT
770
Gly Val Val Leu Ala Tyr Lys Ile Lys Leu Val Arg Val Pro Glu Val
235                 240                 245                 250

GTG ACC GTG TTT ACC ATT GAA AGA AGA GAG GAA CAA AAC CTC AGC ACC
818
Val Thr Val Phe Thr Ile Glu Arg Arg Glu Glu Gln Asn Leu Ser Thr
                255                 260                 265

ATC GCG GAA CGA TGG GTA CAA GTT GCT GAT AAG CTA GAT AGA GAT CTT
866
Ile Ala Glu Arg Trp Val Gln Val Ala Asp Lys Leu Asp Arg Asp Leu
            270                 275                 280

TTC CTT CGA ATG ACC TTT AGT GTC ATA AAC GAT ACC AAC GGT GGA AAG
914
Phe Leu Arg Met Thr Phe Ser Val Ile Asn Asp Thr Asn Gly Gly Lys
        285                 290                 295

ACA GTC CGT GCT ATC TTT CCA ACG TTG TAC CTT GGA AAC TCG AGG AAT
962
Thr Val Arg Ala Ile Phe Pro Thr Leu Tyr Leu Gly Asn Ser Arg Asn
    300                 305                 310

CTT GTT ACA CTT TTG AAT AAA GAT TTC CCC GAG TTA GGG TTG CAA GAA
1010
Leu Val Thr Leu Leu Asn Lys Asp Phe Pro Glu Leu Gly Leu Gln Glu
315                 320                 325                 330

TCG GAT TGT ACT GAA ATG AGT TGG GTT GAG TCT GTG CTT TAC TAC ACG
1058
Ser Asp Cys Thr Glu Met Ser Trp Val Glu Ser Val Leu Tyr Tyr Thr
                335                 340                 345

GGC TTC CCC AGT GGT ACT CCA ACC ACG GCG CTC TTA AGC CGT ACT CCT
1106
Gly Phe Pro Ser Gly Thr Pro Thr Thr Ala Leu Leu Ser Arg Thr Pro
            350                 355                 360

CAA AGA CTC AAC CCA TTC AAG ATC AAA TCC GAT TAT GTG CAA AAT CCT
1154
Gln Arg Leu Asn Pro Phe Lys Ile Lys Ser Asp Tyr Val Gln Asn Pro
        365                 370                 375

ATT TCT AAA CGA CAG TTC GAG TTC ATC TTC GAA AGG CTG AAA GAA CTT
1202
Ile Ser Lys Arg Gln Phe Glu Phe Ile Phe Glu Arg Leu Lys Glu Leu
    380                 385                 390
```

```
GAA AAC CAA ATG TTG GCT TTC AAC CCA TAT GGT GGT AGA ATG AGT GAA
1250
Glu Asn Gln Met Leu Ala Phe Asn Pro Tyr Gly Gly Arg Met Ser Glu
395                 400                 405                 410

ATA TCC GAA TTC GCA AAG CCT TTC CCA CAT AGA TCG GGT AAC ATA GCG
1298
Ile Ser Glu Phe Ala Lys Pro Phe Pro His Arg Ser Gly Asn Ile Ala
                415                 420                 425

AAA ATT CAA TAC GAA GTA AAC TGG GAG GAT CTT AGC GAT GAA GCC GAA
1346
Lys Ile Gln Tyr Glu Val Asn Trp Glu Asp Leu Ser Asp Glu Ala Glu
            430                 435                 440

AAT CGT TAC TTG AAT TTC ACA AGG CTG ATG TAT GAT TAC ATG ACC CCA
1394
Asn Arg Tyr Leu Asn Phe Thr Arg Leu Met Tyr Asp Tyr Met Thr Pro
        445                 450                 455

TTT GTG TCG AAA AAC CCT AGA AAA GCA TTT TTG AAC TAT AGG GAT TTG
1442
Phe Val Ser Lys Asn Pro Arg Lys Ala Phe Leu Asn Tyr Arg Asp Leu
    460                 465                 470

GAT ATT GGT ATC AAC AGC CAT GGC AGG AAT GCT TAT ACT GAA GGA ATG
1490
Asp Ile Gly Ile Asn Ser His Gly Arg Asn Ala Tyr Thr Glu Gly Met
475                 480                 485                 490

GTT TAT GGG CAC AAG TAT TTC AAA GAG ACA AAT TAC AAG AGG CTA GTA
1538
Val Tyr Gly His Lys Tyr Phe Lys Glu Thr Asn Tyr Lys Arg Leu Val
                495                 500                 505

AGT GTG AAG ACT AAA GTT GAT CCT GAC AAC TTC TTT AGG AAT GAG CAA
1586
Ser Val Lys Thr Lys Val Asp Pro Asp Asn Phe Phe Arg Asn Glu Gln
            510                 515                 520

AGC ATC CCA ACT TTG TCA TCT TGA A GAACGTACAT ATATAAATAA
1631
Ser Ile Pro Thr Leu Ser Ser  *
        525                 530

ATACCTTTGT GCATGGTATT TTCAGGGTGT TAAAGTGATA TTCAGATATT TATGATAGAA
1691

TTTTGACTTG TATTTTATAC AATCAAAATT GTATGGTTCT CCGAATTTCT CTTTTTAATT
1751

CTGAAAAATA CATATTAGTA TTGTCAAAAA AAA
1784
```

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 530 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
Met Gln Thr Ser Ile Leu Thr Leu Leu Leu Leu Leu Leu Ser Thr Gln
 1               5                  10                  15

Ser Ser Ala Thr Ser Arg Ser Ile Thr Asp Arg Phe Ile Gln Cys Leu
            20                  25                  30

His Asp Arg Ala Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr
        35                  40                  45

Pro Gly Asn Ser Ser Phe Pro Thr Val Leu Gln Asn Tyr Ile Arg Asn
    50                  55                  60

Leu Arg Phe Asn Glu Thr Thr Thr Pro Lys Pro Phe Leu Ile Ile Thr
65                  70                  75                  80

Ala Glu His Val Ser His Ile Gln Ala Ala Val Val Cys Gly Lys Gln
                85                  90                  95

Asn Arg Leu Leu Leu Lys Thr Arg Ser Gly Gly His Asp Tyr Glu Gly
            100                 105                 110

Leu Ser Tyr Leu Thr Asn Thr Asn Gln Pro Phe Phe Ile Val Asp Met
        115                 120                 125

Phe Asn Leu Arg Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp
    130                 135                 140

Val Gln Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr Arg Ile Ala Glu
145                 150                 155                 160

Lys Ser Asn Lys His Gly Phe Pro Ala Gly Val Cys Pro Thr Val Gly
                165                 170                 175

Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Asn Leu Met Arg Lys
            180                 185                 190

Tyr Gly Leu Ser Val Asp Asn Ile Val Asp Ala Gln Ile Ile Asp Val
        195                 200                 205

Asn Gly Lys Leu Leu Asp Arg Lys Ser Met Gly Glu Asp Leu Phe Trp
    210                 215                 220

Ala Ile Thr Gly Gly Gly Gly Val Ser Phe Gly Val Val Leu Ala Tyr
225                 230                 235                 240
```

Lys Ile Lys Leu Val Arg Val Pro Glu Val Val Thr Val Phe Thr Ile
245 250 255

Glu Arg Arg Glu Glu Gln Asn Leu Ser Thr Ile Ala Glu Arg Trp Val
260 265 270

Gln Val Ala Asp Lys Leu Asp Arg Asp Leu Phe Leu Arg Met Thr Phe
275 280 285

Ser Val Ile Asn Asp Thr Asn Gly Gly Lys Thr Val Arg Ala Ile Phe
290 295 300

Pro Thr Leu Tyr Leu Gly Asn Ser Arg Asn Leu Val Thr Leu Leu Asn
305 310 315 320

Lys Asp Phe Pro Glu Leu Gly Leu Gln Glu Ser Asp Cys Thr Glu Met
325 330 335

Ser Trp Val Glu Ser Val Leu Tyr Tyr Thr Gly Phe Pro Ser Gly Thr
340 345 350

Pro Thr Thr Ala Leu Leu Ser Arg Thr Pro Gln Arg Leu Asn Pro Phe
355 360 365

Lys Ile Lys Ser Asp Tyr Val Gln Asn Pro Ile Ser Lys Arg Gln Phe
370 375 380

Glu Phe Ile Phe Glu Arg Leu Lys Glu Leu Glu Asn Gln Met Leu Ala
385 390 395 400

Phe Asn Pro Tyr Gly Gly Arg Met Ser Glu Ile Ser Glu Phe Ala Lys
405 410 415

Pro Phe Pro His Arg Ser Gly Asn Ile Ala Lys Ile Gln Tyr Glu Val
420 425 430

Asn Trp Glu Asp Leu Ser Asp Glu Ala Glu Asn Arg Tyr Leu Asn Phe
435 440 445

Thr Arg Leu Met Tyr Asp Tyr Met Thr Pro Phe Val Ser Lys Asn Pro
450 455 460

Arg Lys Ala Phe Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn Ser
465 470 475 480

His Gly Arg Asn Ala Tyr Thr Glu Gly Met Val Tyr Gly His Lys Tyr
485 490 495

Phe Lys Glu Thr Asn Tyr Lys Arg Leu Val Ser Val Lys Thr Lys Val
500 505 510

Asp Pro Asp Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Thr Leu Ser
515 520 525

Ser *
530

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CCGCCATGGA GACTTCCATT CTTACTC
27

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GCCGGATCCT CAAGATGACA AAGTTGGGAT GCT
33

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1590 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Helianthus annuus
(B) STRAIN: Zebulon

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1590

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```
ATG GAG ACT TCC ATT CTT ACT CTC CTT CTT CTC TTG CTC TCA ACC CAA
48
Met Glu Thr Ser Ile Leu Thr Leu Leu Leu Leu Leu Ser Thr Gln
  1               5                  10                  15

TCT TCT GCA ACT TCC CGT TCC ATT ACA GAT CGC TTC ATT CAA TGT TTA
96
Ser Ser Ala Thr Ser Arg Ser Ile Thr Asp Arg Phe Ile Gln Cys Leu
            20                  25                  30

CAC GAC CGG GCC GAC CCT TCA TTT CCG ATA ACC GGA GAG GTT TAC ACT
144
His Asp Arg Ala Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr
        35                  40                  45

CCC GGA AAC TCA TCT TTT CCT ACC GTC TTG CAA AAC TAC ATC CGA AAC
192
Pro Gly Asn Ser Ser Phe Pro Thr Val Leu Gln Asn Tyr Ile Arg Asn
    50                  55                  60

CTT CGG TTC AAT GAA ACT ACC ACA CCA AAA CCC TTT TTA ATC ATC ACA
240
Leu Arg Phe Asn Glu Thr Thr Thr Pro Lys Pro Phe Leu Ile Ile Thr
65                  70                  75                  80

GCC GAA CAT GTT TCC CAC ATT CAG GCA GCT GTG GTT TGT GGC AAA CAA
288
Ala Glu His Val Ser His Ile Gln Ala Ala Val Val Cys Gly Lys Gln
                85                  90                  95

AAC CGG TTG CTA CTG AAA ACC AGA AGC GGT GGT CAT GAT TAT GAA GGT
336
Asn Arg Leu Leu Leu Lys Thr Arg Ser Gly Gly His Asp Tyr Glu Gly
            100                 105                 110

CTT TCC TAC CTT ACA AAC ACA AAC CAA CCC TTC TTC ATT GTG GAC ATG
384
Leu Ser Tyr Leu Thr Asn Thr Asn Gln Pro Phe Phe Ile Val Asp Met
        115                 120                 125

TTC AAT TTA AGG TCC ATA AAC GTA GAT ATC GAA CAA GAA ACC GCA TGG
432
Phe Asn Leu Arg Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp
    130                 135                 140
```

GTC CAA GCC GGT GCG ACT CTT GGT GAA GTG TAC TAT CGA ATA GCG GAG
480
Val Gln Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr Arg Ile Ala Glu
145 150 155 160

AAA AGT AAC AAG CAT GGT TTT CCG GCA GGG GTT TGT CCA ACG GTT GGC
528
Lys Ser Asn Lys His Gly Phe Pro Ala Gly Val Cys Pro Thr Val Gly
165 170 175

GTT GGT GGG CAT TTT AGT GGT GGT GGG TAT GGT AAT TTG ATG AGA AAA
576
Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Asn Leu Met Arg Lys
180 185 190

TAT GGT TTG TCG GTT GAT AAT ATT GTT GAT GCT CAA ATA ATA GAT GTG
624
Tyr Gly Leu Ser Val Asp Asn Ile Val Asp Ala Gln Ile Ile Asp Val
195 200 205

AAT GGC AAG CTT TTG GAT CGA AAG AGT ATG GGT GAG GAT TTG TTT TGG
672
Asn Gly Lys Leu Leu Asp Arg Lys Ser Met Gly Glu Asp Leu Phe Trp
210 215 220

GCG ATC ACC GGC GGT GGT GGT GTT AGT TTT GGT GTG GTT CTA GCC TAC
720
Ala Ile Thr Gly Gly Gly Gly Val Ser Phe Gly Val Val Leu Ala Tyr
225 230 235 240

AAA ATC AAA CTA GTT CGT GTT CCG GAG GTT GTG ACC GTG TTT ACC ATT
768
Lys Ile Lys Leu Val Arg Val Pro Glu Val Val Thr Val Phe Thr Ile
245 250 255

GAA AGA AGA GAG GAA CAA AAC CTC AGC ACC ATC GCG GAA CGA TGG GTA
816
Glu Arg Arg Glu Glu Gln Asn Leu Ser Thr Ile Ala Glu Arg Trp Val
260 265 270

CAA GTT GCT GAT AAG CTA GAT AGA GAT CTT TTC CTT CGA ATG ACC TTT
864
Gln Val Ala Asp Lys Leu Asp Arg Asp Leu Phe Leu Arg Met Thr Phe
275 280 285

AGT GTC ATA AAC GAT ACC AAC GGT GGA AAG ACA GTC CGT GCT ATC TTT
912
Ser Val Ile Asn Asp Thr Asn Gly Gly Lys Thr Val Arg Ala Ile Phe
290 295 300

CCA ACG TTG TAC CTT GGA AAC TCG AGG AAT CTT GTT ACA CTT TTG AAT
960
Pro Thr Leu Tyr Leu Gly Asn Ser Arg Asn Leu Val Thr Leu Leu Asn
305 310 315 320

```
AAA GAT TTC CCC GAG TTA GGG TTG CAA GAA TCG GAT TGT ACT GAA ATG
1008
Lys Asp Phe Pro Glu Leu Gly Leu Gln Glu Ser Asp Cys Thr Glu Met
                325                 330                 335

AGT TGG GTT GAG TCT GTG CTT TAC TAC ACG GGC TTC CCC AGT GGT ACT
1056
Ser Trp Val Glu Ser Val Leu Tyr Tyr Thr Gly Phe Pro Ser Gly Thr
            340                 345                 350

CCA ACC ACG GCG CTC TTA AGC CGT ACT CCT CAA AGA CTC AAC CCA TTC
1104
Pro Thr Thr Ala Leu Leu Ser Arg Thr Pro Gln Arg Leu Asn Pro Phe
        355                 360                 365

AAG ATC AAA TCC GAT TAT GTG CAA AAT CCT ATT TCT AAA CGA CAG TTC
1152
Lys Ile Lys Ser Asp Tyr Val Gln Asn Pro Ile Ser Lys Arg Gln Phe
    370                 375                 380

GAG TTC ATC TTC GAA AGG ATG AAA GAA CTT GAA AAC CAA ATG TTG GCG
1200
Glu Phe Ile Phe Glu Arg Met Lys Glu Leu Glu Asn Gln Met Leu Ala
385                 390                 395                 400

TTC AAC CCA TAT GGT GGT AGA ATG AGT GAA ATA TCC GAA TTC GCA AAG
1248
Phe Asn Pro Tyr Gly Gly Arg Met Ser Glu Ile Ser Glu Phe Ala Lys
                405                 410                 415

CCT TTC CCA CAT AGA TCG GGT AAC ATA GCG AAG ATT CAA TAC GAA GTA
1296
Pro Phe Pro His Arg Ser Gly Asn Ile Ala Lys Ile Gln Tyr Glu Val
            420                 425                 430

AAC TGG GAG GAT CTT AGC GAT GAA GCC GAA AAT CGT TAC TTG AAT TTC
1344
Asn Trp Glu Asp Leu Ser Asp Glu Ala Glu Asn Arg Tyr Leu Asn Phe
        435                 440                 445

ACA AGG CTG ATG TAT GAT TAC ATG ACT CCA TTT GTG TCG AAA AAC CCT
1392
Thr Arg Leu Met Tyr Asp Tyr Met Thr Pro Phe Val Ser Lys Asn Pro
    450                 455                 460

AGA GAA GCA TTT TTG AAC TAT AGG GAT TTG GAT ATT GGT ATC AAC AGC
1440
Arg Glu Ala Phe Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn Ser
465                 470                 475                 480

CAT GGC AGG AAT GCT TAT ACT GAA GGA ATG GTT TAT GGG CAC AAA TAT
1488
His Gly Arg Asn Ala Tyr Thr Glu Gly Met Val Tyr Gly His Lys Tyr
                485                 490                 495
```

```
TTC AAA GAG ACA AAT TAC AAG AGG CTA GTA AGT GTG AAG ACT AAA GTT
1536
Phe Lys Glu Thr Asn Tyr Lys Arg Leu Val Ser Val Lys Thr Lys Val
            500                 505                 510

GAT CCT GAC AAC TTC TTT AGG AAT GAG CAA AGC ATC CCA ACT TTG TCA
1584
Asp Pro Asp Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Thr Leu Ser
        515                 520                 525

TCT TG
1590
Ser
    530
```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 529 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```
Met Glu Thr Ser Ile Leu Thr Leu Leu Leu Leu Leu Ser Thr Gln
 1               5                  10                  15

Ser Ser Ala Thr Ser Arg Ser Ile Thr Asp Arg Phe Ile Gln Cys Leu
            20                  25                  30

His Asp Arg Ala Asp Pro Ser Phe Pro Ile Thr Gly Glu Val Tyr Thr
        35                  40                  45

Pro Gly Asn Ser Ser Phe Pro Thr Val Leu Gln Asn Tyr Ile Arg Asn
    50                  55                  60

Leu Arg Phe Asn Glu Thr Thr Thr Pro Lys Pro Phe Leu Ile Ile Thr
65                  70                  75                  80

Ala Glu His Val Ser His Ile Gln Ala Ala Val Val Cys Gly Lys Gln
                85                  90                  95

Asn Arg Leu Leu Leu Lys Thr Arg Ser Gly Gly His Asp Tyr Glu Gly
            100                 105                 110

Leu Ser Tyr Leu Thr Asn Thr Asn Gln Pro Phe Phe Ile Val Asp Met
        115                 120                 125

Phe Asn Leu Arg Ser Ile Asn Val Asp Ile Glu Gln Glu Thr Ala Trp
    130                 135                 140

Val Gln Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr Arg Ile Ala Glu
145                 150                 155                 160
```

Lys Ser Asn Lys His Gly Phe Pro Ala Gly Val Cys Pro Thr Val Gly
165 170 175

Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Asn Leu Met Arg Lys
180 185 190

Tyr Gly Leu Ser Val Asp Asn Ile Val Asp Ala Gln Ile Ile Asp Val
195 200 205

Asn Gly Lys Leu Leu Asp Arg Lys Ser Met Gly Glu Asp Leu Phe Trp
210 215 220

Ala Ile Thr Gly Gly Gly Gly Val Ser Phe Gly Val Val Leu Ala Tyr
225 230 235 240

Lys Ile Lys Leu Val Arg Val Pro Glu Val Val Thr Val Phe Thr Ile
245 250 255

Glu Arg Arg Glu Glu Gln Asn Leu Ser Thr Ile Ala Glu Arg Trp Val
260 265 270

Gln Val Ala Asp Lys Leu Asp Arg Asp Leu Phe Leu Arg Met Thr Phe
275 280 285

Ser Val Ile Asn Asp Thr Asn Gly Gly Lys Thr Val Arg Ala Ile Phe
290 295 300

Pro Thr Leu Tyr Leu Gly Asn Ser Arg Asn Leu Val Thr Leu Leu Asn
305 310 315 320

Lys Asp Phe Pro Glu Leu Gly Leu Gln Glu Ser Asp Cys Thr Glu Met
325 330 335

Ser Trp Val Glu Ser Val Leu Tyr Tyr Thr Gly Phe Pro Ser Gly Thr
340 345 350

Pro Thr Thr Ala Leu Leu Ser Arg Thr Pro Gln Arg Leu Asn Pro Phe
355 360 365

Lys Ile Lys Ser Asp Tyr Val Gln Asn Pro Ile Ser Lys Arg Gln Phe
370 375 380

Glu Phe Ile Phe Glu Arg Met Lys Glu Leu Glu Asn Gln Met Leu Ala
385 390 395 400

Phe Asn Pro Tyr Gly Gly Arg Met Ser Glu Ile Ser Glu Phe Ala Lys
405 410 415

Pro Phe Pro His Arg Ser Gly Asn Ile Ala Lys Ile Gln Tyr Glu Val
420 425 430

Asn Trp Glu Asp Leu Ser Asp Glu Ala Glu Asn Arg Tyr Leu Asn Phe
435 440 445

Thr Arg Leu Met Tyr Asp Tyr Met Thr Pro Phe Val Ser Lys Asn Pro
450 455 460

Arg Glu Ala Phe Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn Ser
465 470 475 480

His Gly Arg Asn Ala Tyr Thr Glu Gly Met Val Tyr Gly His Lys Tyr
485 490 495

Phe Lys Glu Thr Asn Tyr Lys Arg Leu Val Ser Val Lys Thr Lys Val
500 505 510

Asp Pro Asp Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Thr Leu Ser
515 520 525

Ser

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..350

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GAGAAACTCG GAGACTTTCA CACAATGCCT AACCTCAAAC TCCGACCCCA AACATCCCAT
60

CTCCCCCGCT ATCTTCTTCT CCGGAAATGG CTCCTACTCC TCCGTATTAC AAGCCAACAT
120

CCGTAACCTC CGCTTCAACA CCACCTCAAC TCCGAAACCC TTCCTCATAA TCGCCGCAAC
180

ACATGAATCC CATGTGCAAG CCGCGATTAC TTGCGGGAAA CGCCACAACC TTCAGATGAA
240

AATCAGAAGT GGAGGCCACG ACTACGATGG CTTGTCATAC GTTACATACT CTGGCAAACC
300

GTTCTTCGTC CTCGACATGT TTAACCTCCG TTCGGTGGAT GTCGACGTGG
350

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..278

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGCATGGATC TCCGCCGGAG CGACTCTCGG AGAGGTTTAT TATCGGATTT GGGAGAAAAG
60

CAGAGTCCAT GGATTCCCCG CCGGAGTTTG ACCGACGGTT GGTGTTGGTG GGCATTTAAG
120

CGGCGGTGGT TACGGTAACA TGGTGAGGAA GTTTGGATTA TCTGTGGATT ACGTTGAGGA
180

TGCCAAGATC GTCGATGTAA ACNGTCGGGT TTTAGATCGG AAAGCAATGG GTGAGGATCT
240

GTTCTGGGCG ATTACCGGTG GAGGAGGAGG TAGCGTAC
278

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..345

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TGGACATATT AGCGGAGGAG GATTCGGTAC AATAATGAGG AAATACGGTT TAGCGTCTGA
60

TAACGTTGTG GACGCACGTT TGATGGATGT AAATGGGAAA ACTCTTGACC GGAAAACGAT
120

GGGAGAGGAT TTGTTTTGGG CGCTTAGAGG CGGTGGAGCT GCGAGTTTTG GCGTTGTCTT
180

GTCGTGGAAG GTTAAGCTTG CTAGGGTTCC TGAAAAGGTA ACTTGTTTCA TAAGTCAACA
240

TCCGATGGGA CCTAGCATGA ACAAGCTTGT TCATAGATGG CAATCCATAG GATCAAGANN
300

GCTAGACGAA GATTTATTCA TCAGAGTCAA TATTGACAAC AGTCT
345

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 695 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..695

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GTTCGTTAAA ACCTATCCTN NANGGGCNAA AGNATATCAA AGNTTGNTTA NGNAACCCAA
60

NATTTCTGAA CTGGCCNCCT TCGGTGGTAT ATGNCNAAAN CCCTTGAATC TGCGNANCCN
120

ATTCCGCATA GAAACGGAAC CCTCTTCAAG ATTCTCTATT TACNCGAACT GNCTAGANNG
180

AATGACAAGA CATCGAGTAG NAAAATCAAC TGGATCAAAG AGATATACAA TTACATGGCG
240

CCTTATGTCT CAAGCAATCC AAGACAAGCA TATGTGAACT ACAGAGATCT AGACTTCGGA
300

CAGAACAAGA ACAACGCAAA GGTTAACTTC ATTGAAGCTA AAATCTGGGG ACCTAAGTAC
360

TTCAAAGGCA ATTTTGACAG ATTGGTGAAG ATTAAAACCA AGGTTGATCC AGAGAACTTC
420

TTCAGGCACG AGCAGAGTAT CCCACCTATG CCCTACTAGA AGCTAGGTTC ATGAAACCAA
480

TAACATTATC AAAAATAAGR ATAAATGRTA ATTGTATACA ACATGATTCG KCTTTCTTTA
540

TTTCAGACAA TGTGGACACT ACTCTAAANT AAAAWGTCNA TTTACCTTAA AAAAAAAATA
600

ATCCCCNNTA ANANAAAANT GGGGGGGCCN TTTTTGGGGN TCCCGGTTTT NGGACGGGGN
660

GCTTTNGGGG GGCTTGGNNT TTTTTTNGGN GCCCC
695

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..495

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TCTGTTTTNA GGCAGAGCAG AGGAAGTTGT TGCTTTGCTT GGTAAGGAGT TTCCTGAATT
60

NAGTTTAAAG AAGGAGAACT GTTCGGAGAT GACTTGGTTT CAGTCAGCTT TATGGTGGGA
120

TAATCGTGTT AACCCTACTC ANATTGATCC WAAAGTGTTT CTCGATCGGA ATCTTGATAG
180

AGCGAATTTC GGAAAGAGGA AATCGGATTA CGTTGCGAGT AAGATTCCTA GAGATGGGAT
240

TAAGYCTTTT TCCAAGARGA TGMCTGACCT GGGGAAAAYC GGGCTTGTTT TTAAWCCGTA
300

TGGTGGGAAA ATGGCGGAGG TTACGGTTAA CGCGACGCCG TTTCCNCACC GAAGCAAGCT
360

TTTTAAGATT CAGTACTCGG TGACTTNGCA AGAAAACTCT NTCGAGATAG AGAAAGGGTT
420

TCTTGAATCA GGCTAACGTC CTTATAGGTT CATGACCGGG TTTTTNAGCA AGANCCCTGG
480

AATNCTTACT TNAAT
495

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..204

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AAATTAAAAC AAATCAATGT TGATATTGAA TCCAATAGTG CTTGGTTTCA ACCTGGTGCT
60

ACGCTTGGTG AGCTTTACTA CAGAATTNCA GAGAAGAGCA AAATCCATGG ATTTCCNGCG
120

GGTTTNTNCA CAAGCNTAGG CATAGGTGGG TATATNANAG GCGGTGGATA CGGTACCTTG
180

ATGAGGAAGT ATGGTCTTNC GGGA
204

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 491 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..491

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GAGATTTCTC GAGCAAGATA CTCCACTGAT GATCTTTGAG CCATTGGGTG GGAAAATCAG
60

CAAGATTTCA GAAACAGAAT CTCCATATCC ACACAGAAGA GGTAATCTGT ATAATATACA
120

GTACATGGTG AAATGGAAAG TGAATGANGT CGAGGAGATG AACAAACATG TCAGGTGGAT
180

GAGATCGTTA CACGATTACA TGACTCCGTA TGTTTCTAAA TCGCCGAGAG GAGCTTATTT
240

GANTTACAGA GATCTTGATT TGGGCTCGAC CAAAGGGATT AACACGGGTT TCGGAGATGC
300

AAGGAAATGG NNGGGTGAGN CTTTTTTCAA AGGTAATTTC CAAGGGGTTA GGTTTTGGTT
360

AAAGGGGAGG TTTNNCCCAN CAAATTTTTT TTCAGGANCC GGCCANGNTT TTCCCCCCCC
420

TNTTTTTNGG NCCCCAATCN AAANCCCCGT TTTAAAAGGG GGGCCATTTC NTTTTTTNCA
480

NNTTAAAAGG G
491

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..407

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATTTGTTCGT GAGGTTAACT TTGACTTTAG TCAACGGTAC GAAGCCTGGT GAGAATACGG
60

TTTTAGCGAC TTTCATTGGG ATGTATTTAG GCCGGTCGGA TAAGCTGTTG ACCGTNATGA
120

ACCGGGATTT CCCGGAGTTG AAGCTGAAGA AAACCGATTN TACCGAGATG AGATGGATCG
180

ATTCGGTTCT GTTTTGGGAC GATTATCCGG TTGGTACACC GACTTCTGTG CTACTAAATC
240

CGCTAGTCGC AAAAAAGTTG TTCATGAAAC GAAAATCGGA CTACGTGAAG CGTCTNATTT
300

TCGAGAACCC GATCTCNNGT TTGATACTCA AGAAATTTGT AGAGGTTNNG AAAGTTAAAA
360

TNAATTTGGA TCCGCATTNN GGNANNNATG GTGAAACCCC NNGTTNT
407

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..360

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ACGGCGTCGT ATTGGCCTAC AAAATAAACC TTGTTGAAGT CCCAGAAAAC GTCACCGTTT 60

TCAGAATCTC CCGGACGTTA GAACAAAATG CGACGGATAT CATTCACCGG TGGCAACAAG 120

TTGCACCGAA GCTTCCCGAC GAGCTTTTCA TAAGANCAGT CATTGACGTA NAAACGGCAC 180

TGTTTCATNN CTCAAAAGAC CGTCAGACAA CATTCATAGC AATGTTTCTA GGAGACACGN 240

CAACTCTACT GTCGATATTA AACCGGAGAT TCCCAGAATT GGGTTTGGTC CGGTCTGACT 300

GTACCGNAAC AAGCNNTTGG ATCCAATCTG TGCTATTTTT GGGACAAATA TCCCAGGTTG 360

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 427 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..427

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TCTTCACTGT CACCAAAACG TTAGAACAAG ACGCAAGATT GAAGACTATT TCTAAGTGGC
60

AACAAATTTC ATCCAAGATT ATTGAAGAGA TACACATCCG AGTGGTACTC AGAGCAGCTG
120

GAAATGATGG AAACAAGACT GTGACAATGA CCTACCTAGG TCAGTTTCTT GGCGAGAAAG
180

GCACCTTGCT GAAGGTTATG GAGAAGGCTT TTCCAGAACT AGGGTTAACT CAAAAGGATT
240

GTACTGAAAT GAGCTGGATT GAAGCCGCCC TTTTCCATGG TGGRTTTCCA ACAGGKTCTC
300

CTATTGAAAT TTTGCTTMAG CTCAAGTCGC CTYTAGGAAA AGRTTWCTTC AAAGCAACGK
360

CGGATTTCGT TAAAGAACCT WTTCCTGTGA TAGGGCTCAA AGGAATATTC AAAAGATTGA
420

TTGAAGG
427

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..437

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GTTGTACTAT CATNGAAGAT TAAGTTAGTC GATGTTCCGT CCACGGTCAC CGNGTTTAAA
60

GTCCAGAAAC ATNAGGAGAA AGAGGCCGTT AGGNTCATCA ACAAGTGGCA GTATGTTGCG
120

GATAAGGTCC CTGAAGATCT TTTCATCAGC GCAACGTTGG NGAGATCAAA CGGAAACTCT
180

GTGCAGGCTT TGTTTACTGG ACTCTATCTT GGNCCGGTGA ATAATNTCTT GGCCTTGATG
240

GAAGAAAAGT TTCCAGANTT AGGTCTTGAT ATCCAAGNCT GCACAGAGAT GAGTTGGGCT
300

GAATCTGCAC TCTGGTNTNC TGNTTTCNCT AAAGGAGAGN CTCCTTGGGT GTTCCNCGCG
360

GATCGGNAGC GGNCAATTTN TGGNCTTTCA AGGGGAAAGN CGGCTTTTTN CAAGAACCCG
420

NTACCCGGGG TTCAATT
437

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..441

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GCGGACCCTA TAGATCANNA TGTGCTACTG ANAGAAGAGG AAGCCAAGAA CAAGCCGGAG
60

ACAGATAAAT ATCTGAAATG GGNCGATANC GTTTACGAAT TTATGACNCC ATATGTTTCG
120

AAATCTCCAA GAGGAGCTTA TGTCAATTTC AAGGATATGG ATTTGGGTAT GTATCTTGGA
180

AAGAAGAAGA CAAAGTACGA GGAAGGAAAG AGTTGGGGAG TGAAGTATTT CAAGAACAAT
240

TTCGAGAGAT TGGTGAGAGT GAAGACTAGG GTTGATCCAA CAGATTTCTT CTGCGATGAA
300

CAGAGCATTC CTCTGGTGAA CAAAGTTACC TGAAGATATC ATTTGAAGTT TTTTATTAGT
360

CCCTTTTCTC TGTGAAATCA TCTGTGCGTG TTGAATATTA TGCGTCAAGT GTGTAACTTA
420

TGTGTGTGAT TGTGAATTGT G
441

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 502 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CTGGCTTAAC ACAACGTCGT TTTGGGCCAA TTACCCGGCG GGTACACCCA AGAGCATCCT
60

TCTAGATAGG CCTCCGACGA ATTCAGTGTC ATTTAAGAGT AAATCGGATT TTGTCAAAAA
120

ACCAATACCC AAAAAAGGTT TAGAGAAGCT TTGGAAGACA ATGTTTAAAT TCAACAGTAG
180

CGTCTCGTTG CAATTCAACC CTTACGGTGG AGTGATGGAC CGGATTCCGG CAACGGCCAC
240

CGCTTTTCCT CATCGGAAAG GAAACTTGTT CAAGGTTCAA TACNCTACGA TGTGGTTTGA
300

CGCAAACGCC ACACAGAGTA GCCNGGCTAT GATGAATGAG CTTTTTGAGG TGGCGGGACC
360

GTACGTGNGT CAAGTAAACC CGAGANANGG CTTCCTTTAA NTTCAGAGNC CATCGNTNTT
420

NGGAGCAANN CCAAGTGGGG GGGNCCAACC GGGGGNTNAA ANCNNAGNTC TTNGGGGGCC
480

CAGAATTTCC TTNGGGGAAT TT
502

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 400 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..400

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

NGGGAATTGC NCGAGGNAAG TTGTACCCAA TTCCTGGACC ACCATTGGTT TCCCAAGAAN
60

CCCGAGACAA CCGTTTTTCA ATNACCGTGA TGTTGATTTG GGTATTAATT CTCATAATGG
120

TAAAATCAGT AGTTATGTGG AAGGTAAACG TTACGGGAAG AAGTATTTCG CAGGTAATTT
180

CGAGAGATTG GTGAAGATTA AGACGAGAGT TGATAGTGGT AATTTCTTTA GGAACGAACA
240

GAGTATTCCT GTGTTACCAT AAGTGTATTT ATTTGATTAT TGGTTAGTGA AATTTGTTGT
300

TGTATAATGA TTATATGTCG TATTTTTATT TATTATTAGT AATTTATAAA GTTTGATATT
360

AAATACAAAT AGTATAATAA GATAGTTTCT TTTAGTAAAA
400

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..383

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CAACTCTAAT GGGAACACCT ACTTCGATCG AATGTCGATG GGGGAAGAGC TTTTCTGGGC
60

GGTTCGAGGA GGTGGAGCCG CGAGTTTCGG CATCGTGATG GGATACAAAA TCCGGTTGGT
120

TCCGGTTCCG GAGAAAGTTA CGGTTTTTAG CGTCGGAAAA ACCGTCGGAG AAGGAGCCGT
180

TGATCTTATA ATGAAGTGGC AGAACTTCTC TCATAGTACG GNTCGGAATT TNTTTGTGAA
240

GCTGANTTTT GANTTTAGTC AACGGTGCAA AGCCGGGTGA AAAAAAGGTT TTAGNGNCTT
300

TCANTTTGGN TGNAANCTTG GGGGTTTTAT NAGAACGGTT AACCGGGATT NANCCCGNGT
360

TTTCCCGGGG TTAAAACCTT NGG
383

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..354

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

ATCAATGTTC TTACTAAACG TACACGAGCA TCGTTGGCTT TCAAGGCTAA ATCTGATTTT 60

NTTCAAGAAC CGATNCCTAA AACCGCGATT TCGAAGCTTT GGAGACGGTT GCAAGAACCG 120

GAAGCAGAGC ATGCTCAGCT AATTTNCACN CCATTTGGTG GTAAAATGAG TNAGATTGCA 180

GATTACGAAA CACCATTTCC GCATAGGAAG GGGAATATAT ATNAGATTCA GTACTTGAAT 240

TACTGGAGAG GAGACGTGAA AGAGAAGTAT ATTGAGATNG GTGGAGGAGA GTTTACGGTT 300

GNTATNAGTA AGTTTTTTGG CGAAGTNTNC CNAGAGGNGN CTTNNTNTAA ACCT 354

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidposis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..403

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TTTTTTAGTA CACTAATAAT CAAATGGAAT GAGAAATGAA GCCACAAAAG TATCTGCAAT
60

CAAAATATCC TGCTATCTCC ATCTCAAGCT CTCAATAGTA TCCTCTCCGA AAGTGAAATC
120

AACATTTCAA ACTCTATTTC TTGGTGGAAT CGATAGACTG ATTCCTCTGA TGAACCAGAA
180

GTTTCCGGAA CTCGGCTTAC GATCTCAAGA CTGTTCGGAA ATGAGCTGGA TCGAATCGAT
240

AATGTTCTTC AACTGGAGAT CAGGACAGCC GTTAGAGATT TTGCTCAACA GAGACCTAAG
300

GATTCGAGGA TCAGTATTTC AAAGCAAAGT CAGGATTATG GTTCAAAAAC CCGTTCCTGA
360

AAACGTTTTT CGAAGAGGTA TCCAAGGGGT TTCTCGAGCA AGT
403

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..260

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GAGATGAGTT GGATTAANTC TGTACTCTGG TTTGCTGATT TCCCTAAAGG AGAATCTCTT
60

NGTGTTCTCA CGAATCGTAA GCGTACATCT CTATCTTTNA AAGGCAAAGA TGATTTTATC
120

CAAGAACCGA TACCCGAGGC TGCAATTNAA GAGATATGGA GGCGATTAGA AGCCCCCNAG
180

GCTCGGCTTG GAAAGATCAT ATTAACTCCA TTTGGTGGGA AAATNAGTGA AATGGCAGAG
240

TACGTANCAC CATTCCCACA
260

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 605 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..605

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CTCTTGCATA TTCGCTGCAA GGATGGGAAA TTCAAAACCA CTCCCTACAA TTTTTTGTAT
60

TATAGTTTCA GTCTTGTATT TTTAATTCTA TTGCATAACA CCAACTTCTT CATCAGCCTC
120

CATCCAAGAT CAATTCATAA ACTGTGTCAA AAGAAACACA CATGTTTCTT TTCCACTCGA
180

GAAAACGTTA TTCACCCCTG CGAAAAACGT CTCTTTGTTC AACCAAGTCC TTGANTCGAC
240

GGCTCAAAAT CTCCAGTTCT TGGCAAAATC CATGCCTAAA CCGGGRTTCA TATTCAGACC
300

GATTCACCAG TCTCAAGTCC AAGSTTCCAT CATTTGTTCA AMGRAACTCG GGNTTCATTT
360

TNGTGTTTGA NGTGGCGGTC ACGATTTTCG AGGCCTTTGT NTTTATGTTT CACGGTTTGA
420

AAAAACCGTT TATATTACTC GGCCTGTCAA ANTTGNANNC AAAATCANAT GTTGGATATT
480

GNATTCCAAA TAGGTNCTTG GGGTNAACCT GGTGGCTANC GTTTGGTGAG CTTTTACTTT
540

CAAGAATTTG CANGNGGANG TGCAAAGATT CCATGGGATT TCCCGGGGGG TTTNTTGCAC
600

AATGT
605

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 464 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..464

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AACACAAAAC TCTTCCATTT GGCTTCTCTC TTGCATATTC GTTGCAAGGA TGGGAAATTC
60

AAAACCACTC CCTACAATTN CTTGTATTAT CGTTTCAGTC TTGTATTTTN NATTCTATTG
120

CATAACACCA ACTTCTTCAT CAGCCTCCAT CCAAGNTCAA TTCATAAACT GTGTCAAAAG
180

GAACACACAT GTTTCTTTTC CACTCGAGNA AACGGTATTC ACTCCTGCGG AAAACGGCTC
240

TNTTATTCAA CGGGTCCNTG AATCGACGGG TCAAAATCTC CAGTTCTTGG NAAAATCCAT
300

GNCTAAACCG GGGTTCATAT TCAGGCCGGT TCACCAGTCT CAAGTCCAAG NTTCCATCAT
360

TTGTTCAAAG GAACTCGGGA TTCATTTCCG CGNTAGAAGT GGCGGGCANN GGTTTCGGGG
420

CCTGTCTNTT GNTTANGGGN AGGAAAACCG GTTNTATTNC TCGG
464

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 386 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..386

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TCGGGAGCCC ANGNTAAATT ANNTGAAAAT GGGGNCGNAT ANCCGTTTAC NGAATTTTAT
60

GACNCCCAAT ATGTTTCGAA ATCTCAAAGA NNGGGANCTT ATGTCAATTT CAAGGATATG
120

GATTTGGGTA TGTATCTTGG AAAGNAGAAG ACAAAGTACG AGGAAGGAAA GAGTTGGGGA
180

GTGAAGTATT TCAAGAACAA TTTCGAGAGA TTGGTGAGAG TGAAGACTAG GGTTGATCCN
240

ACAGATTTCN TCTGCGATGA ACAGAGCATT CCTCTGGTGN ACAAAGTTAC CTGAAGATAT
300

CATTTGAAGT TTTTTATTAG TCCCTTTTCT CTGTGAAATC ATCTGTGCGT GTTGAATANT
360

ATGCGTCAAG TGTGTAACTT ATGTGT
386

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 377 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..377

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

TACCATAGGG AGGTGGTGNA AGATTTTGTA TGTAGNCTTA GGGGAAGGCG AGTAGTATGG
60

TGGTGGTGGG GAGCTGTAAA CGTATGGTGG TGGTGGAGAT TTGTATGTGG GCTGGTTAAC
120

TTCATTGAAG CTAAAATCTG GGGACCTAAG TACTTCAAAG GCAATTTTGA CAGATTGGTG
180

AAGATTAAAA CCAAGGTTGA TCCAGAGAAC TTCTTCAGGC ACGAGCAGAG TATCCCACCT
240

ATGCCCTACT AGAAGCTAGG TTCATGAAAC CAATAACATT ATCAAAAATA AGAATAAATG
300

ATAATTGTAT ACAACATGAT TCGTCTTTCT TTATTTCAGA CAATGTGGAC ACTACTCTAA
360

ATAAAATGTC ATTTACC
377

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

GAGCTGTGGA TATGGTCACA AATGGCAATC GGTTGGTCCG AAAACTGATC CGAATCTTTT 60

TATGAGAATN TTGATTCAAC CAGTGACGAG GAAGAAGGTA AAGACTGTGA GAGCTTCTNT 120

GGTTGCCCTN TTTTNAGGCN AGACAGATGA AGTTTTTGCT TTCCTTAGTA AGGAGTTTCC 180

TGAATTGGGT TTAAAGAAGG AGAATTNTTC GGAGATGACT TGGTTTCANT CTGCTTTATG 240

GTGGGACAAT CGTCTTAATG CTACTCAGGT TGATCCTAAA GTNTTTCTTG ATCGGAATCT 300

CGATACCTCG AGTTTCGGTA AGAGGAAATC GGATTACGTC GCGACT 346

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..261

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

ATGGGGTGAG ACTTATTTCA AAGGTAATTT CAAGAGATTA GGTTTGGTTA AAGGGAAGNT
60

TGATCCAACA AATTTCTTCA GGAACGAACA GAGTATTCCT CCTCTGTTTT GAGTCCTCAA
120

TACAAAACCA GATATAAAAG ATGTCATTTC ATTTTTTCAA TTATAATAGA TAATGTAACT
180

TTCTGCTACA ATTGTAAAAG TGAGATGTAC CCAATACGGT TTAAGCGGAC CGAGAATAGT
240

CAATTCAAAG ACCAAATTCT G
261

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..478

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GCTCAAAGGA CTAACCATGA AAACTTCCTC AAGTGTCTCT CTCACCGANT CAACGAGGAC
60

GACTCAAGAN TTATACACAC ATCAAAAGAT CCTTCGTATT TNTCAATCTT GATTTCTTCC
120

ATACAAAATC CAAGTTTCTC TGTTCTCGAA ACACCTAAAC CGGTTTCAAT CATCACTCCG
180

GTTCAAGCCA CCGATGTTCA ATCTACGNTT AAATNCGCAC GGNCTTCACG GGTATACACA
240

ATCAGGGCTA GGAGTGGTNG TCATGACTAC GGAGGTTTAT CTTTACATTG GCTTAAAAAN
300

CANNCCGTTC GTTNNTCATT GATTTNNAGA AATCTTCCGG GCTTATTTAA CATNTAAGAT
360

GTTTGATAAN CCGGNNCCNG TTTGGGGTTC AAATCCCGGT GGCTTACAAA NTTNGGGGGA
420

ATTGTNCCTA TGAGGTTTGG AAAATTAANG CAAAATNTTT TGGGCCTTCC CGGCCGGT
478

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 579 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: ecotype Columbia

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..579

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGCCGTTAGG ATCATCAAGA AATGGCAATA TGCTGCAGAT AAGGTTCCTG ATGATCTTTT
60

CATTAGGACA ACATTGGAGA GATCAAACAA GAACGCAGTA CACGCTTTGT TCACTGGACT
120

ATATATTGGT CCGGTGAACA ATCTATTGGC GTTGATGGAA GAAAAGTTTC CGGAACTAGG
180

TCTTGAGAAA GAAGGTTGTG AAGAGATGAG TTGGATTGAG TCTGTACTCT GGTTTGCTGA
240

TTTCCCTAAA GGAGAATCTC TTGGTGTTCT CACGAATCGT GAGCGTACAT CTCTATCTTT
300

CAAAGGCAAA GATGATTTTG TCCAAGAACC GATACCCGAG GCTGCAATTC AAGAGATATG
360

GAGGCGATTA GAAGCCCCCG AGGCTCGGCT TGGAAAGATC ATATTAACTC CATTTGGGTG
420

NGGNAAAATG AGTGAAATGG CAGAGNCCGA ACCACCAATT CCCACANNCG AGGGAGGGGA
480

ACCCCTNTGN GGNTCAGAAT GTGGTTCCTG GNNNNNAAGN GGGNGCCAGN ACCAANCCGG
540

GNCNGTAAAN CNTGNAATGG GCCNAACCCG TNCCGGATT
579

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa
(B) STRAIN: Nipponbare, subsp. japonica
(D) DEVELOPMENTAL STAGE: etiolated shoot (8 days old)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..252

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TGTCCTGGAA GGTCCGCCTC GTGCAGGTTN CGACGACGGT GACGGTGTTC GTCGTCGGGA
60

GGAACGTCGA CCAGGGCGCC GCNGACGTCG TCGCCAGATG GCAAGACGTC GCGCCGAGCC
120

TCCCTCCCGA GCTCACCATA CGGGTGATCG TNCGAGGGCA GCGCGCCACG TTCCAGTCGC
180

TGTACCTCGG CTCGTGCGCC GACCTGGTGC CGACGATGAG CAGCATGTTC CCGGAGCTCG
240

GGATGACGAT TG
252

CLAIMS

1. An isolated protein which has antifungal activity, preferably anti-Oomycete activity, more preferably anti-*Phytophthora* and/or anti-*Pythium* activity and which is obtainable from a plant source, encoded by the nucleotide sequence as shown in SEQ ID NO: 15 or SEQ ID NO: 18 or parts or muteins therefrom.

2. An isolated protein according to claim 1, characterised in that it is obtainable from sunflower plants.

3. An isolated protein having reticuline oxidase activity, characterised in that it has antifungal activity, preferably anti-Oomycete activity and more preferably anti-*Phytophthora* and/or anti-*Pythium* activity.

4. An isolated protein, characterised in that it comprises one or more of the peptides selected from the group consisting of:
(a) amino acids 1 to 25 of SEQ ID NO: 1,
(b) amino acids 1 to 25 of SEQ ID NO: 2,
(c) amino acids 1 to 118 of SEQ ID NO: 6,
(d) amino acids 1 to 529 of SEQ ID NO: 16, or a part of said sequence having antifungal activity,
(e) amino acids 1 to 529 of SEQ ID NO: 20, or a part of said sequence having antifungal activity,
as well as muteins thereof which have antifungal activity.

5. An antifungal protein comprising an amino acid sequence characterised in that it is capable of being encoded by the open reading frame represented by SEQ ID NO: 15, or by part of said open reading frame.

6. An antifungal protein comprising an amino acid sequence characterised in that it is capable of being encoded by the open reading frame represented by SEQ ID NO: 19, or by part of said open reading frame.

7. An antifungal protein comprising an amino acid sequence characterised in that it is capable of being encoded by one of the open reading frames represented by SEQ ID NO's: 21 - 47.

8. An isolated DNA sequence comprising an open reading frame capable of encoding a protein according to any of the claims 1 to 7, and DNA capable of hybridising therewith under stringent conditions.

9. An isolated DNA sequence according to claim 8, characterised in that it comprises the nucleotide sequence depicted in SEQ ID NO: 5.

10. An isolated DNA sequence according to claim 8, characterised in that it comprises the nucleotide sequence depicted in SEQ ID NO: 15.

11. An isolated DNA sequence according to claim 8, characterised in that it comprises the nucleotide sequence depicted in SEQ ID NO: 19.

12. An isolated DNA sequence according to claim 8, characterized in that it comprises one of the nucleotide sequences depicted in SEQ ID NO's: 21 - 47.

13. A chimeric DNA sequence comprising a DNA sequence according to any of claims 8 to 12.

14. A chimeric DNA sequence according to claim 13, further comprising a transcriptional initiation region and, optionally, a transcriptional termination region, so linked to said open reading frame as to enable the chimeric DNA to be transcribed in a living host cell when present therein, thereby producing RNA which comprises said open reading frame.

15. A chimeric DNA sequence according to claim 14, wherein the RNA comprising said open reading frame is capable of being translated into protein in said host cell, when present therein, thereby producing said protein.

16. A chimeric DNA sequence according to any one of claims 13 to 15 which is a replicon, preferably pMOG1144 or pMOG1180.

17. A chimeric DNA sequence according to claim 16 which is a vector.

18. A vector according to claim 17, which is a binary vector, preferably pMOG1144 or pMOG1180.

19. A host cell comprising a replicon according to claim 16 and which is capable of maintaining said replicon once present therein.

20. A host cell comprising a vector according to claim 17 or 18 and which is capable of maintaining said vector once present therein.

21. A host cell stably incorporating in its genome a chimeric DNA sequence according to claim 13 or 14.

22. A host cell according to claim 21 which is a plant cell, said vector being a non-integrative viral vector.

23. A host cell according to claim 21 which is a plant cell.

24. A plant or a plant part comprising at least one plant cell according to claim 22 or 23.

25. A plant or a plant part consisting essentially of plant cells according to claim 23.

26. A plant according to claim 25, characterised in that said chimeric DNA is expressed in at least a number of the plant's cells causing the said antifungal protein to be produced therein.

27. A method for the production of a protein with antifungal activity, preferably anti-anti-Oomycete activity, more preferably anti-*Phytophthora* activity and/or anti-*Pythium* activity, characterised in that a host cell according to claim 20 to 23 is grown under conditions allowing the said protein to be produced by said host cells.

28. A method according to claim 27, further comprising the step of recovering the protein from the host cells.

29. Use of a protein according to any one of claims 1 to 7 for retarding fungal growth, preferably Oomycete growth and more preferably the growth of *Phytophthora sp.* and/or *Pythium sp.*

30. Use of reticuline oxidase for retarding fungal growth.

31. The use according to claim 29 or 30, characterised in that spores of the said fungus are caused to be contacted with said protein.

32. A method of retarding the growth of the a fungus, preferably an Oomycete, more preferably *Phytophthora* or *Pythium* on plant leaves, characterised in that hyphae thereof, or spores thereof, are caused to be contacted with a protein produced from a host cell according to claim 20 or 21, or from a cell of a plant according to claim 26.

33. A method for obtaining plants with reduced susceptibility to fungi, preferbly Oomycetes, more preferably *Phytophthora* or *Pythium*, comprising the steps of

(a) introducing into ancestor cells which are susceptible of regeneration into a whole plant,

- a chimeric DNA sequence comprising an open reading frame capable of encoding a protein according to any of claims 1-7, said open reading frame being operatively linked to a transcriptional and translational region and, optionally, a transcriptional termination region, allowing the said protein to be produced in a plant cell that is susceptible to infection by said fungus and
- a chimeric DNA sequence capable of encoding a plant selectable marker allowing selection of transformed ancestor cells when said selectable marker is present therein, and

(b) regenerating said ancestor cells into a plant under conditions favouring ancestor cells which have the said selectable marker, and

(c) identifying a plant which produces a protein according to claim 1-7, thereby reducing the susceptibility of said plant to infection by said fungus.

34. The method according to claim 33, characterised in that step (a) is performed using an *Agrobacterium tumefaciens* strain capable of T-DNA transfer to plant cells and which harbours a binary vector, and wherein step (b) is performed in the presence of an antibiotic favouring cells which have a neomycin phosphotransferase.

35. An antifungal composition comprising a protein according to any one of claims 1 to 7, and a suitable carrier.

36. An antibody capable of recognising a protein according to any one of claims 1 to 7.

37. A nucleic acid sequence obtainable from a gene encoding a protein according to any one of claims 1 to 7, having tissue-specific and/or developmental specific transcriptional regulatory activity in a plant.

38. A nucleic acid sequence according to claim 37, which is obtainable from the region upstream of the translational initiation site of said gene.

39. A nucleic acid sequence according to claim 38, which has at least 1000 nucleotides of said region upstream of the translational initiation site of said gene.

40. Use of a nucleic acid sequence according to any one of claims 37 to 39 for making a plant expressible gene construct.

ABSTRACT

The present invention provides an isolated protein obtainable from a plant source which has antifungal activity, specifically anti-*Phytophthora* activity and/or anti-*Pythium* activity and a molecular weight of about 59 ± 5 kDa as judged by SDS PAGE-electrophoresis, an isolated DNA sequence comprising an open reading frame capable of encoding a protein according to the invention, preferably characterised in that it comprises an open reading frame which is capable of encoding a protein depicted in SEQ ID NO: 16 or muteins thereof, and DNA capable of hybridising therewith under stringent conditions. The invention further comprises plants incorporating chimeric DNA capable of encoding a protein according to the invention, and wherein the protein is expressed. Also methods are provided for combating fungi, especially *Phytophthora* and *Pythium* species, using a protein or a host cell capable of producing the protein.

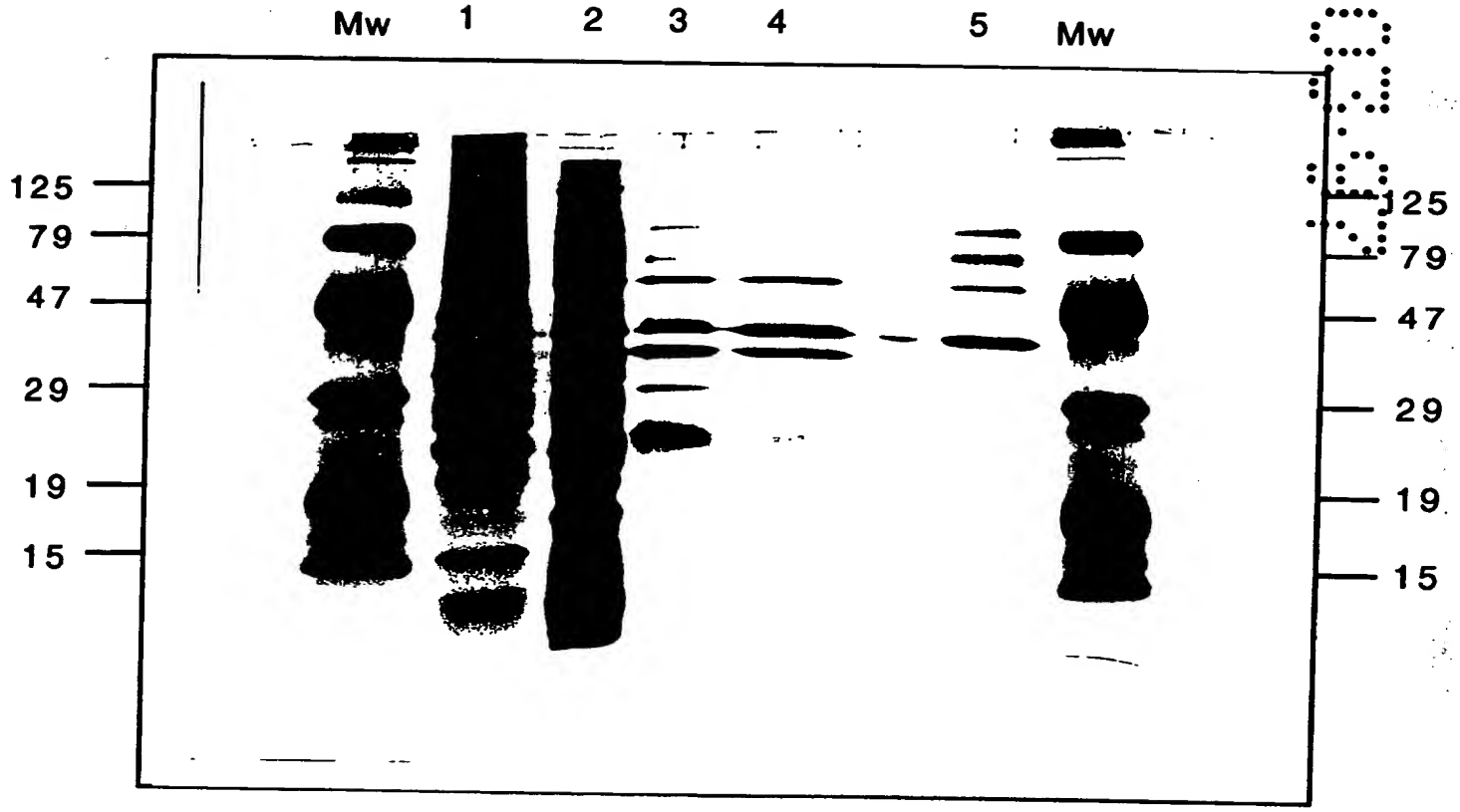
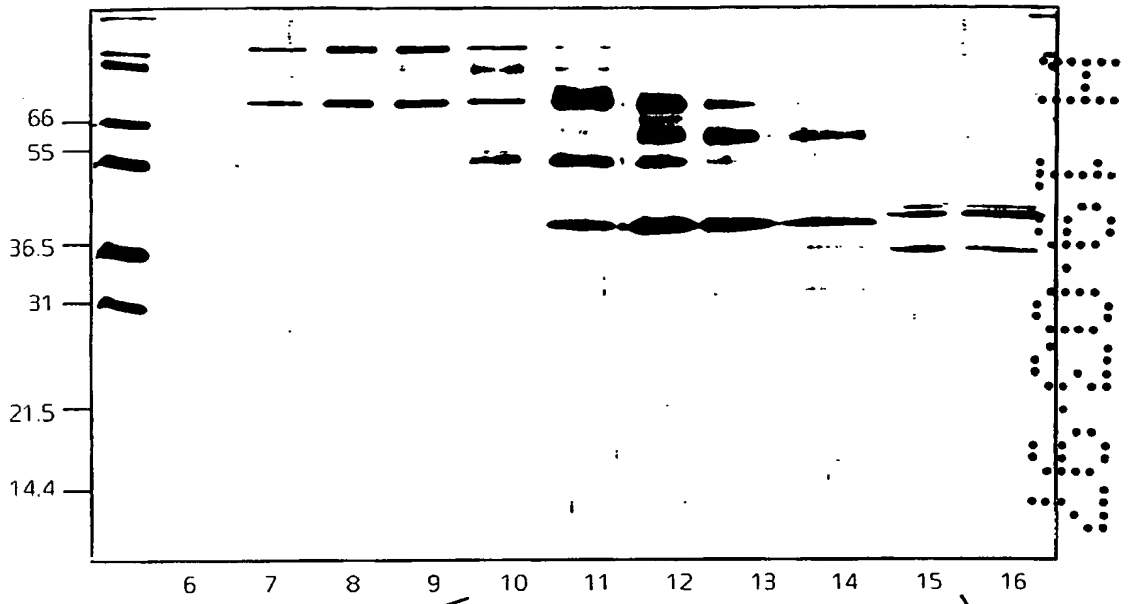


Fig. 1


66
55
36.5
31
21.5
14.4
6
7
8
9
10
11
12
13
14
15
16

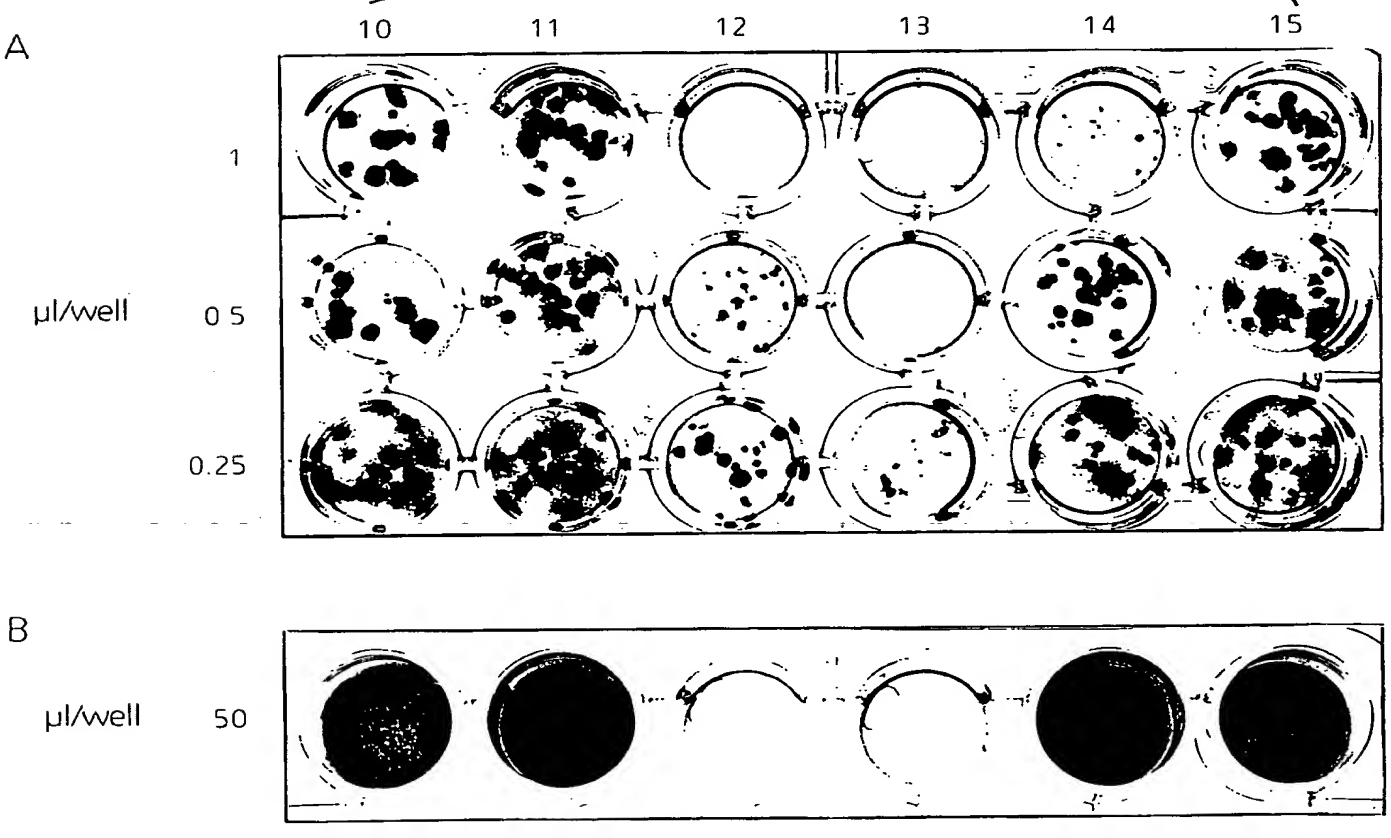

A
10
11
12
13
14
15
1
µl/well
0.5
0.25
B
µl/well
50

Fig. 2

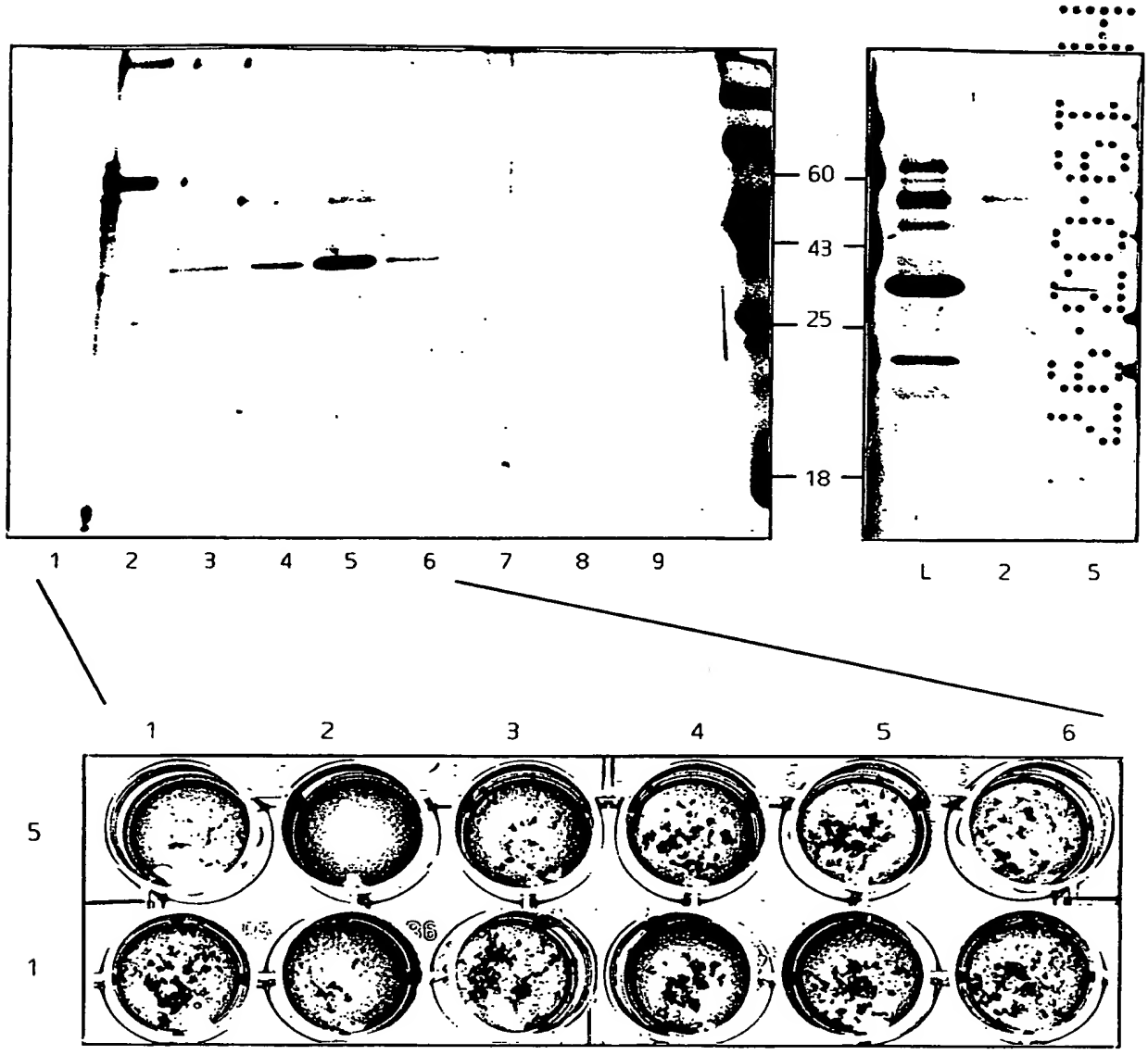


Fig. 3

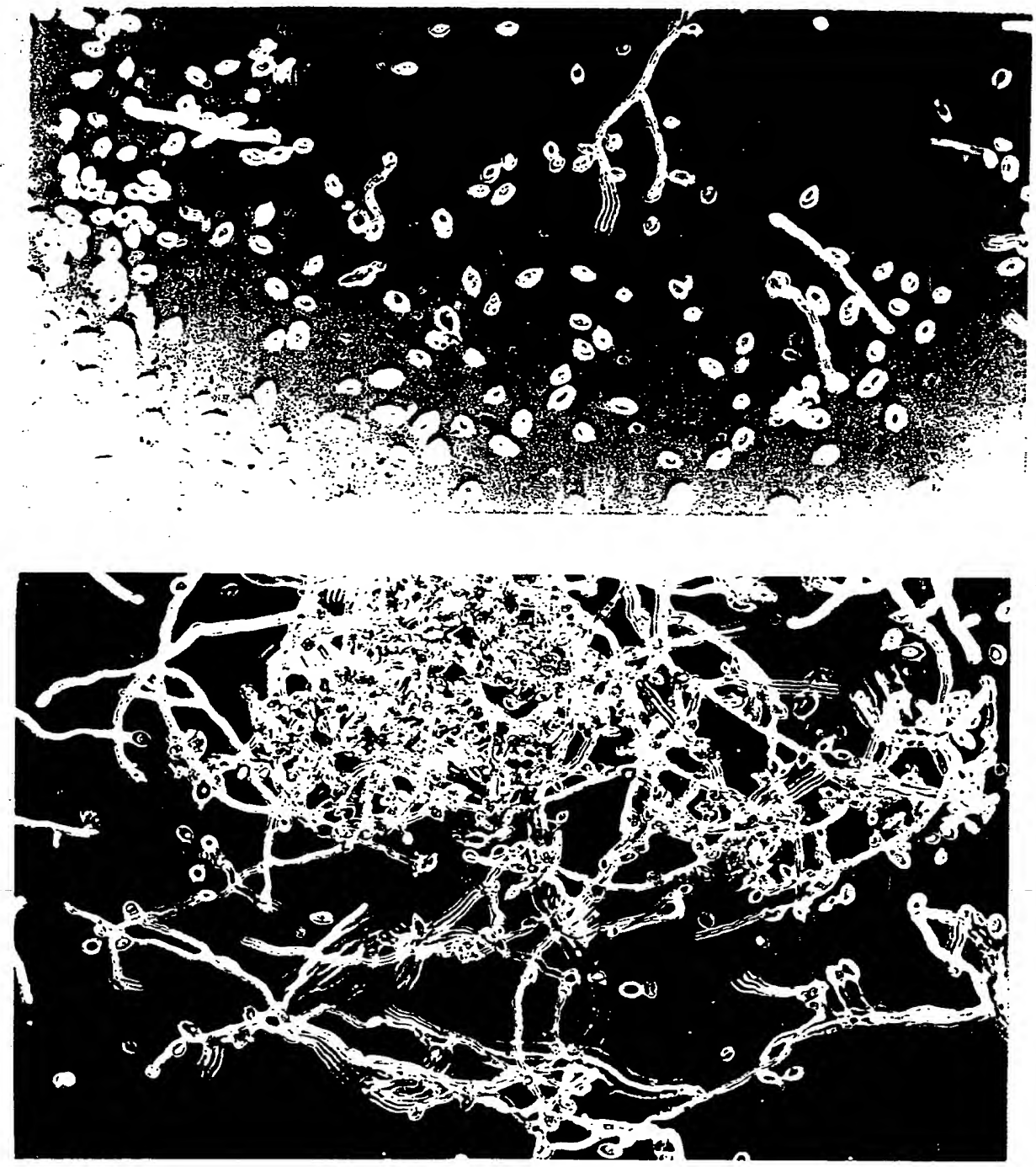


Fig. 4

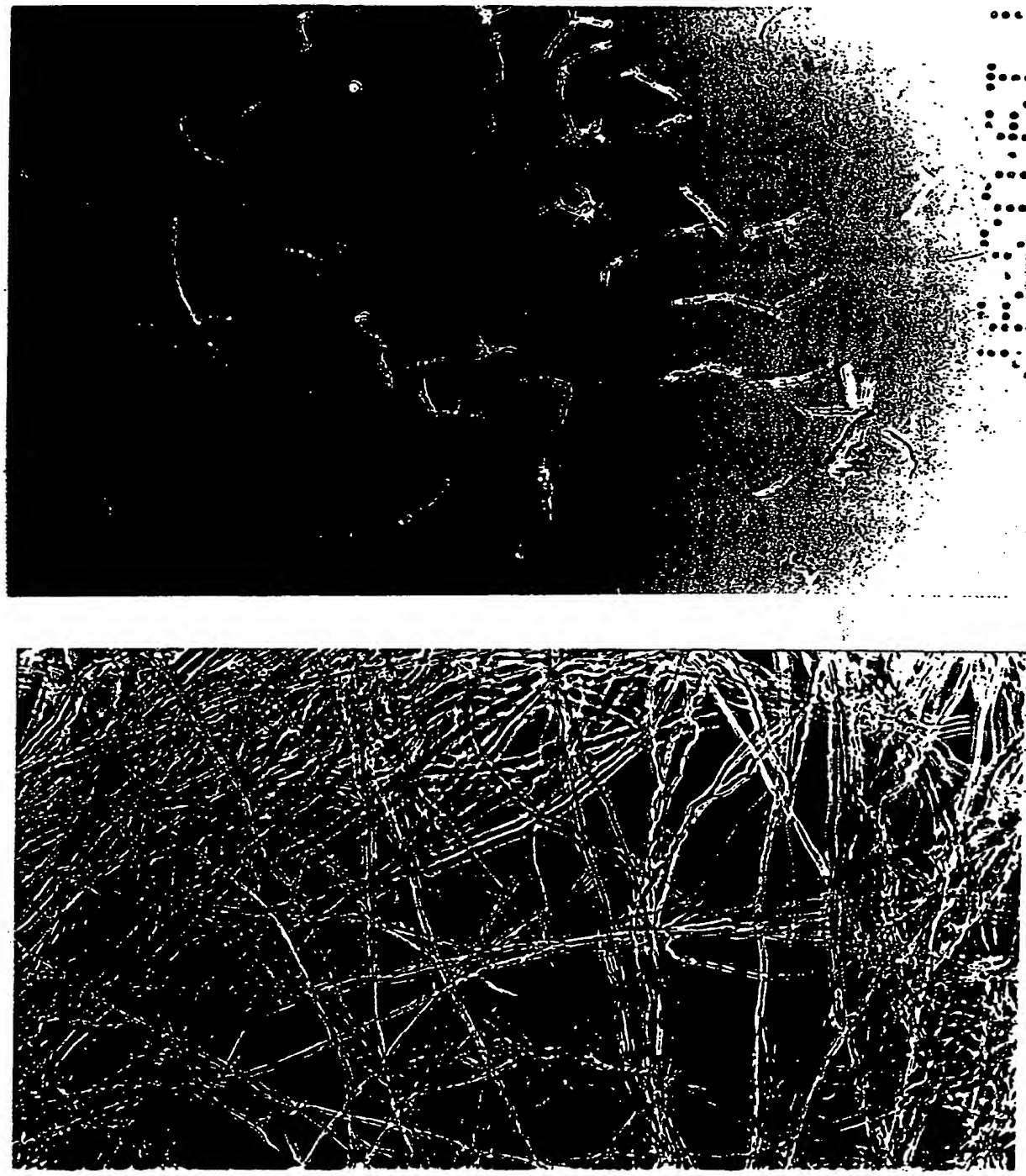


Fig. 5